A Compendium of Position Papers from the Workshop on Architectures for Free Space Digital Optical Computing

Holiday Inn Chateau Vail
Vail, Colorado
January 28th-30th 1991

Organizing committee: Alan Craig (AFOSR)
Miles Murdocca (Rutgers University)
Michael Prise (AT&T Bell Labs)

Administrator: Priscilla Rasmussen (Rutgers University)

Sponsoring organizations: AFOSR, ONR, DARPA

Correspondence with regard to this compendium can be addressed to Miles Murdocca, Department of Computer Science, Rutgers University, Hill Center, New Brunswick, NJ, 08903, (908) 932-2001 (Department secretary), murdocca@cs.rutgers.edu.

Additional copies of this compendium may be obtained by contacting Alan Craig or Miles Murdocca. Addresses for Craig and Murdocca are listed on the last page of the compendium.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL 25 Jan 91 TO 24 Nov 91

4. TITLE AND SUBTITLE
WORKSHOP ON FREE-SPACE OPTICAL COMPUTING ARCHITECTURES

5. FUNDING NUMBERS
2305/B1

6. AUTHOR(S)
Dr Miles Murdocca

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rutgers - The State University of NJ
Pistcataway NJ 08854-1089

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NE
Bldg 410
Bolling AFB Washington DC 20332-6448
Dr. Alan E. Craig

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFOSR-91-0334

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

On January 28-30, 1991, a workshop on architectures for free-space digital optical computing was held at the Holiday Inn Chateau Vail in Vail, Colorado. The workshop was initiated by Alan Craig of the Air Force Office of Scientific Research and was organized with Miles Murdocca (Rutgers University) and Michael Prise (AT&T Bell Labs). The purpose to the workshop was to bring together a panel of distinguished contributors to the field, identify current directions and discuss the future of the field. The emphasis of the workshop was on overall system architectures. Since systems depend on devices and optics, a number of contributors in these areas were invited both to provide a perspective on implementations and to learn what additional progress is necessary in order to implement systems that are either competitive or that complement future digital electronic systems.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Preface

On January 28-30, 1991, a workshop on architectures for free-space digital optical computing was held at the Holiday Inn Chateau Vail in Vail, Colorado. The workshop was initiated by Alan Craig of the Air Force Office of Scientific Research and was organized with Miles Murdocca (Rutgers University) and Michael Prise (AT&T Bell Labs). The purpose of the workshop was to bring together a panel of distinguished contributors to the field, identify current directions and discuss the future of the field. The emphasis of the workshop was on overall system architectures. Since systems depend on devices and optics, a number of contributors in these areas were invited both to provide a perspective on implementations and to learn what additional progress is necessary in order to implement systems that are either competitive or that complement future digital electronic systems.

On the second night of the workshop, three breakout sessions were held covering the topics "Speculative ideas," "What could or should work?" and "What we know that does not work." The chairs for each of these sessions summarized the discussions for their groups in the next few pages. A final compendium of position papers follows, as well as a list of attendees.

In reading the summaries of the breakout sessions bear in mind that these views reflect the opinions of participants on the second night of the workshop and may not reflect the views of the entire community or even the views of the same participants today.

Speculative ideas

Group Chair: Miles Murdocca

• SuperTuring computing.

• Electronic computing with free space electronic interconnects, as in an electron beam sweeping a CRT.

• Optical dataflow machines. (For those unfamiliar with dataflow architectures, this is a fairly untested area of computing, and I guess that's why this is in the speculative section. - Chair)

• Implementation of optical neural networks.

• Reconfigurable interconnects and steerable interconnects, the speculative aspects being the speed of operation, and what influence these ideas have on computing.

• Superconducting optical neural computers (No comment from the chair.)

• Modal representation (as opposed to amplitude representation) of a binary (or N-ary) signal.

• Fast, parallel access, low-energy, multidimensional memory. (From the Chair: Electronic memories are typically fast like static RAM, or dense like dynamic RAM, and we live with these differences quite well in electronic technologies, so maybe the fast and low energy criteria should serve as goals rather than requirements. The parallel access and multidimensional features would be useful in conventional computers today.)

• Volumetric displays.

• Optical database machines. Discussion suggested optics may be good for correlations and associations, and that there may be a new capability for representing spatial objects.

• Free-space storage. Quote from Alan Huang: "Free space is something we have plenty of."

• Optical fuzzy logic.

• Reversible logic. This is still a way-out topic for electrical and mechanical systems as well. (See *What we know that does not work* section. - Ed.)

• Self-learning optical neural networks, as suggested by an avid supporter of optical neural networks.

• All-optical free-space computing that is fast and runs at low energy. This idea runs counter to current trends.

• Low-cost irregular interconnects. This topic relates to work that favors regular interconnects. It was suggested that an observer may have the mistaken impression that regular interconnects are better than irregular interconnects on the basis of interconnection power, when in fact, regular interconnects are a special case of the more general class of irregular interconnects, and it is costs vs. benefits of each that should be argued.

• Global nonlinearities - One neuron behaves as many, via wavelength multiplexing.

What could or should work?

Group Chair: Mike Prise

• Modifications to existing computer architectures

• Carefully designed optical clocks

• Opto-electronic isolators in wafer scale integration

• Wide short length array data links with low threshold lasers

- Optical Lego blocks (1 micron tolerance) - connectorized optics
- WDM backplanes
- Fast light modulators on silicon
- GaAs integrated circuits with lasers
- Optical digital computers (no comment from the Editor.)
- Database machines (see *Speculative ideas*. - Ed.)
- Parallel access optical disks
- Useful random holographic optical interconnect
- Medium scale crossbars
- GaAs ICs with optoelectronics
- Software tools for interconnects
- Free-space board-to-board connectors
- Dataflow machines - with packet switches (see *Speculative ideas*. - Ed.)
- Parallel access optical memory
- Point to point link

(For sale by 2002)

- Optical computers
- Free-space optically interconnected multiprocessors
- Analog optical correlator - portable, small, cascaded multiple channels
- Optical switch for telecommunications of size 1024x1024
- Fine grain optical processor

• Parallel DRAM

• Holographic interconnects

What we know that does not work (except perhaps in special applications)

Group Chair: Alan Craig

• Von Neumann designs, *i.e.* single processor machines do not fit in well with optical processing technology.

• Reversible, dissipationless logic was discounted on the basis of no current need, and the requirement of consuming energy in drawing any conclusion or making any decision. The quantum mechanical computer was denigrated in the same breath by the same cynic. (Each of these futuristic ideas was defended by a protagonist - see the *Speculative ideas* section).

• Vector-matrix processors with serial input (better architectures are available in silicon); digital systolics are better, and systolics are better digital. In the same vein, no locally connected architectures appear to be suitable for optics since electronic technologies support locally connected processing elements quite well. An objection to pipeline latency was countered by the observation that throughput can still be high - this issue was unresolved.

• Associative processors, particularly those that do not require an exact match (dynamic range and noise problems in discrimination exist). Content addressable memory was argued to be a niche variety of an associative processor with viability - good for database processing.

• Cellular automata has only niche applications, such as modelling of fluid dynamics, and perhaps as comprising self organizing systems. Not much use for numerics, and hard to program.

• Pure symbolic substitution cannot be practically configured for numerics. (Symbolic substitution is functionally complete, but requires greater fan-outs, more space, and longer latency than a more conventional approach. - Ed.)

• No optical binary correlators for general purpose processing (electronics are more proficient).

• Concern was expressed regarding the latency of array interconnects, particularly in fine grain systems with feedback. This is paramount to declaiming global interconnects, particularly irregular interconnects, are not useful except perhaps in large grain architectures.

• Waveguide optics has short term uses for functions more akin to signal processing, and may enhance nonlinearities and some switches, but are generally antithetical to computing architectures. Path layout is difficult despite non-interfering crossings, partially due to losses at tight bends.

(However, later presentation by R. Linke on slab waveguide broadcast interconnects refuted some of this conclusion.)

• Coherence may be useful only for a large number of frequency channels (N > 10,000 suggested.)

• Resolution inadequacy severely limits prospects for an all-optical computer. Many electronic devices can populate the area (approximately a few microns square) of an optical emitter/detector.

• Optics should not feel tied to residue arithmetic; pipelining the carries may sustain the life of residue. (Early optical computing was forced into residue arithmetic due to the lack of an appropriate feedback mechanism. Recall that residue arithmetic can be performed in a single step, thus the attraction to this method for one or two-stage optical systems. - Ed.)

• Don't use AND gates. (This addresses the difficulty of distinguishing various levels of light as opposed to just the presence or absence of light as for OR and NOR. - Ed.)

• Photorefractives are suspect for interconnects. Even with careful exposure scheduling, gratings compete for dynamic range, and crosstalk persists. Also, these are only 2-terminal devices. Uses in self-aliging systems may be found, but in reconfigurable interconnects less likely. Materials development efforts continue.

• Late entry; shadow-casting logic seems quite primitive.

Digital Optical Computing: Honeywell SRC's Perspective

Julian Bristow
Honeywell, Inc.
Systems and Research Center
10701 Lyndale Ave. South
Bloomington, MN 55420-5601

Proponents of digital optical computing have predicted orders of magnitude improvement in system performance over electronic systems. To date, the promise remains unfulfilled. Optical computers will only gain acceptance with potential users and systems integrators when the system performance exceeds that available from conventional technology by at least an order of magnitude.

Current efforts in digital optical computing build on paradigms employed for electronic systems. For general purpose computing therefore, improvements in system performance can only be demonstrated if each and every element of the system offers improvement over its electronic counterpart. Thus not only must logic elements be developed with some improvement in parallelism, speed or other metric, but for example a complete hierarchy of long and short term memory must be provided. Particularly for shorter term memory, gaps exist in the technology. In addition to the memory elements themselves, decode and control mechanisms must be provided.

Specific problems may always be found which circumvent the issues. Indeed, departure from the aims of general purpose computing and focussing on special purpose processors is likely to yield the first demonstrations of digital optical computing. Many real applications of computing involve embedded processors operating on only a narrow range of problems. Development of the system elements with the required performance will require selection and analysis of specific special purpose applications. Restriction to special purpose processors will however keep the cost of the technology high, and limit its acceptance even for special applications. Therefore, the greater goal of developing a general purpose computer remains an attractive one, and will depend on demonstration of improved materials, devices and packaging technology.

Key practical issues must be addressed in consideration of implementation of any architecture. Neglect of such vital considerations as packaging and alignment, even at this early stage, will result in limited acceptance of the technology. The issue of packaging refers not only to the three dimensional alignment of components or modules (itself a difficult task, since conventional packaging techniques are two dimensional), but also to thermal management. Most optical devices demonstrated to date are inefficient in terms of power, yet are sensitive to changes in operating temperature. These properties limit the number of processing elements which may be integrated in a given space for a given system speed. Demonstration of viable systems will require that techniques be developed to remove unwanted heat from the three-dimensional system. Issues of thermal management are especially pertinent for military systems, in which optical computing may ultimately have high payoff. Here the wide variations in operating temperature, together with the density of components which will be required to demonstrate improvements over electronic processors indicate that radically improved packaging technology will be required. These limitations will continue to apply unless conventional paradigms are abandoned.

MULTIFUNCTIONAL OPTICAL/DIGITAL FREE SPACE PROCESSORS

DAVID CASASENT
Carnegie Mellon University
Center for Excellence in Optical Data Processing
Department of Electrical and Computer Engineering
Pittsburgh, PA 15213

Workshop on Architectures for Free Space Digital Optical Computing
Colorado, January 1991

Optical/digital processors (defined as systems capable of general-purpose logic/numeric processing to floating point accuracy) may be possible (given device advancements). We feel that such systems should be used to perform high level (numeric not logic) functions for the multitude of matrix-vector operations/applications that exist. Thus, we feel that only an optical numeric array processor is a viable optical/digital processor (specifically, it should perform additions, multiplications, and vector inner products).

Within the above constraints, there are 3 issues that are needed to define the optical realization of such a system: the data source/sink (parallel page-addressed optical memory), the number representation (a new formulation of MSD, since it avoids carries beyond one bit location), and the architecture (a cascaded correlator, since such architectures exist in well engineered form). Figure 1 shows the block diagram of our system. With optically-addressed ferroelectric liquid crystal page composers (input and output SLMs) and with a space/frequency multiplexed filter bank (containing the recognition/substitution rules) as in Figure 2 (where the laser diode activated provides access to the proper set of 9 recognition/substitution filters), the performance of such a system can exceed 10^{10} OPS (as we will show).

We next consider the fact that an optical processor should be general purpose or multifunctional. We specifically consider the use of such an optical/digital processor for image processing functions. The image processing functions we consider are all low level vision operations (morphological operations) and a hierarchical/inference processor (for pattern recognition and certified intelligence functions).

The result is a multifunctional optical processor for computer vision and a viable general-purpose optical processor.

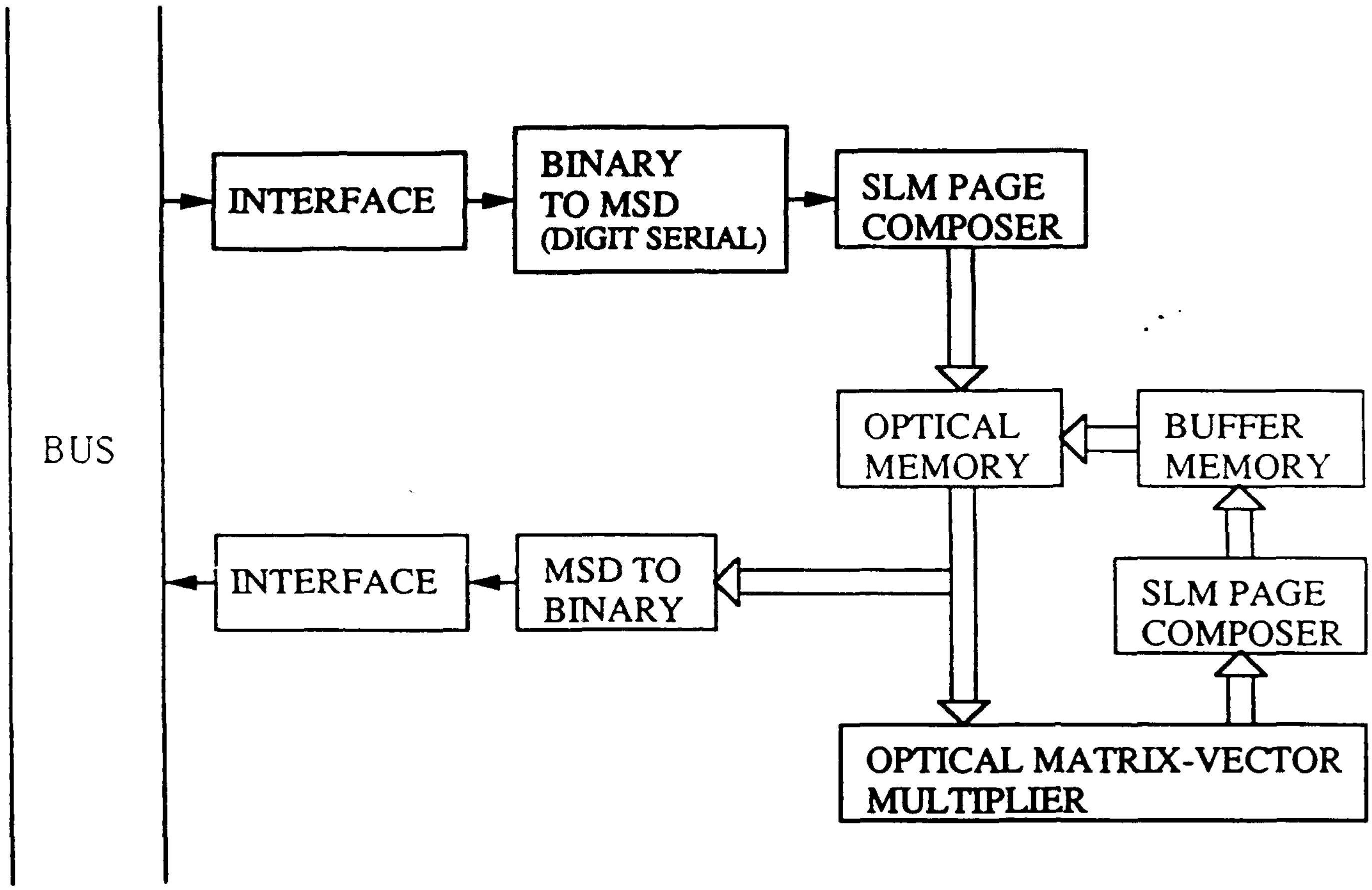


FIGURE 1: Optical/digital numeric processor concept.

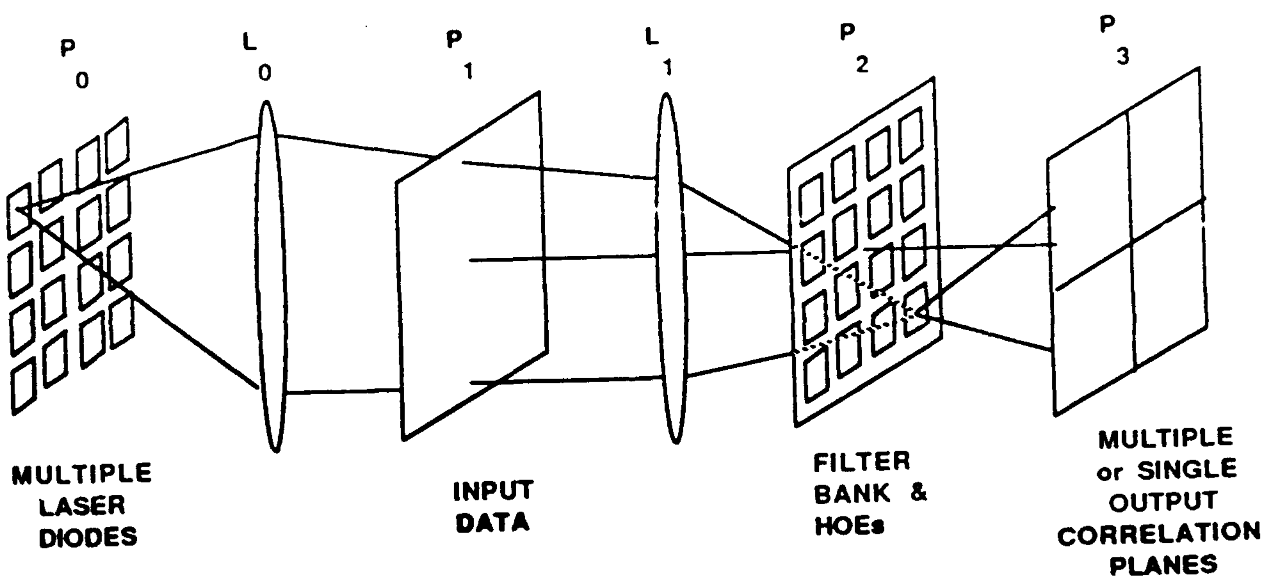


FIGURE 2: MIMD LD (laser diode) addressed processor.

IS PRESENTLY MORE RESEARCH ON OPTICAL ARCHITECTURES NEEDED ?

P. Chavel, Ph. Lalanne, J. Taboury,
Institut d'Optique (CNRS), B.P. 147, 91403 Orsay cedex, France

Optical logic demonstrators like for example those made at ATT Bell Labs with S-SEED appear to be primarily limited by the technology of the new devices needed, and in particular their energy requirements. The perspective of millions of gates on a few square centimeters of active devices is currently hardly realistic, whereas this performance is typical with CMOS technology with femtojoule switching energies. It is a commonplace to say that the largest expectations for optics are in the domain of interconnects for massive parallelism and high data flow rates. The general purpose processor under development at Opticomp, that is expected to reach clock frequencies over 100 MHz and uses a combination of electronics, acoustics and optics but no nonlinear optics, relies on optics mostly for propagating many beams through free space. In this paper, we examine the role of optical interconnects in three (somewhat loosely defined) kinds of processors classified by their degree of architecture novelty compared to an electronic machine.

I - Optics in conventional silicon computers :

We first consider the case where optics could be introduced with minimal disturbance to the designer of an electronic system. Communications between chips and even more so communications between boards are limited by so-called impedance matching conditions that in turn limit the bandwidth of the whole system. This limitation is related to the capacity of lines, that itself is determined by the necessity to minimize electromagnetic interference. Therefore, optics should be first useable to increase clock frequency through the reduction of interchannel modulation rejection. This can be achieved at minimal disturbance to the architecture using fiber optics links and falls outside the scope of this workshop firstly because it is not free space optics and secondly because it hardly poses optical architecture problems. It may, however, pave the way to our second category of processors.

II - Massive optical i/o for silicon based processors :

This category still relies mostly on silicon chips placed on boards ; its existence is justified by two arguments :
a) the chips and boards interconnects do not only limit the system clock frequency, but also set severe constraint on circuit and system architecture because of the limited number of input output channels to/from each chip or board. 3-D optical input output may change the life of designers by providing a number of i/o channels of the order of the number of computing sites ;
b) analyses of energy balances and computational throughput have been published by several authors, indicating that optical interconnects at the gate to gate level appear to be a promising alternative for communication lengths in the range of a several millimeters to centimeters, and this would be an important issue at least for W.S.I..

The use of optics to provide massive point to point connexions in such systems would imply to find viable solutions to several open questions :
• space needed to place optical interconnect components between boards and between chips ;
• ruggedness and alignment procedures and tolerances of the optics with the electronics ;
• use of intermediate components, probably new compound semiconductor devices, for the output, since silicon can be used to detect light, but not to modulate or emit light ; because no light has to be emitted or modulated at the chip level is one reason for the interest of the optical clock distribution problem, that may be considered either as in the context of increasing clock frequency (section I) or in that of massively distributing an optical signal (this section).

If solutions are found to these issues in terms of optoelectronic and passive optical devices and optical systems, then the computer designers may well accept to go through the effort of adapting their architectures to accomodate some optics.

III - Specific architectures for optics :

This last category includes all optical digital computing architecture concepts that have shown up in the last few years and that use not only massive optical interconnects through free space, but the combination of the optical signals to form weighted sums as well. These include "cellular processors" with optical interconnects, i.e. optical neural processors and optical cellular automata/symbolic substitution processors. There, the operations performed by optics are fan-in, fan-out, matrix-vector multiplication and convolution (or correlation) and can be interpreted in terms of binary pattern recognition.

Some groups (such as UCSD) investigate the combination of silicon and optical modulators to validate the concept of such processors with the important additional advantage that the considerable computing power of silicon is incorporated ; other approaches rely strictly on novel nonlinear optoelectronic devices. Both seem good to us as soon as performances are improved from one generation to the next, but the first may well outperform pure silicon in a shorter term.

One open issue may be the kind of applications that are suitable for these machines : while some seem to advocate general purpose machine and it is known that such processors may easily have the power of Turing machines, it has apparently not been shown that they are potentially a good solution for making Turing machines ; on the other hand, specialized applications such as associative machines, low level image processing or (our favorite example) statistical physics (the "lattice gas automaton") may be interesting in principle but not respond to large needs.

Another open issue, in our opinion the most important, is that of optical systems as opposed to optical architectures. Most demonstrators up to now are bulky, difficult to align, and perform extremely modest operations not only because of device limitations but also because of system poverty : for example, can an optical neural processor without learning be very interesting, can an optical symbolic substitution processor working on two-bit, non-programmable symbols be useful (we have guilty for some of these ourselves in the past). Shadow casting approaches and spatial frequency approaches do not seem easy to make rugged and compact with a good space-bandwidth product, maybe something may be expected from lenslet arrays and hologram arrays but this remains to be investigated.

Conclusions

What is meant by digital optical computing architectures ?
• is it the definition of a number of operations that could be performed optically in future competitive processors,
• or is it the definition of a combination of such operations into a machine such as the optical symbolic substitution processor or an optical neural network processor, using principles such as polarization encoding of data, wavelenght encoding, dark-pixel recognition of binary patterns, dual-rail encoding,
• or is it the implementation of such a processor with primitive equipment ?
In these three cases, we think that not much more research is needed in the domain for now.

However, we do think that much more is to be done on the subject and that many necessary ideas are still missing if to work on "digital optical computing architectures" means
• develop optical systems concepts needed to considerably improve the compactness and space-bandwidth product of the demonstrators,
• develop and make the appropriate passive components, given active components that exist or are expected to come soon,
• combine all these into record breaking processors.

Architectures for free-space digital optical computing

T. J. Cloonan

The use of free-space optics for digital computing has been on the chalk boards of researchers for more than a decade now, and although quite a few promising research efforts have produced very interesting results on paper and a few small lab experiments, the actual number of results that have proven themselves to be useful for practical systems is depressingly low. Most of the architectures that have been proposed in the literature to date (including several by this author) have been techically infeasible or economically impractical. This does not imply that these architectural proposals are without value, because they have served as guidelines for the embryonic optical technologies (devices, lasers, optics, and opto-mechanics) that are developing along with the system architectures, and as these technologies mature, some of the architectures may become realizable. Nevertheless, system architects in the field of free-space digital optics are shouldering a heavy load, because they must try to design sytem architectures using technologies that don't yet exist, and their designs must be able to perform better (in terms of functionality and cost) than the electronic-based systems that will appear in the field X years from now when the optical technologies have matured. Thus, optical system architects must aim at a moving (perhaps accelerating is a better word?) target with weapons whose performance is still undetermined. If optical system architects have nothing else, they should at least have a good guess as to where they think they should be X years from now, and they should also have a path defined which might get them there. Thus, the beginning of a new decade of research marks a good time to re-examine our directions and re-evaluate our goals to determine suitable paths for the future.

In the view of this author, the path toward an "all-optical computer" seems to be an impractical path to follow at this time, because the switching energies of the available optical logic devices require too much laser power. To decrease this switching energy, many device researchers are looking into logic gates with small amounts of localized electronic gain. This is the beginning step toward a "smart pixel." An obvious extension of this approach would lead to optical detectors driving electronic logic whose outputs drive optical modulators or sources. The localized gate-level interconnections are electronic, while the longer-distance interconnections between functional units are optical. Electronic technologies do some things incredibly well, and limited processing in a localized area on an electronic chip is one of those things. Localized processing of this sort does not require long interconnection lengths, so signals can be transported between the logic gates at very high bandwidths without requiring power-hungry terminating resistors on the transmission lines. In addition, many of the interconnections within these localized processing units are fairly random, and the optics required to perform these random interconnections become fairly complicated and expensive. Thus, electronics for localized processsing and optics for longer-distance interconnections seems to be a good compromise solution, and "optical purists" who ignore the benefits of electronics are giving up a large amount of processing power that can be obtained by letting optics augment the functionality of electronics. (Depending on the their area of expertise, some researchers might view it as letting electronics augment the functionality of optics, but its all the same).

If we assume that a hybrid solution of optics and electronics might yield useful results, the next question

that must be answered is how much of the system is optical and how much of it is electronic. Another way to ask this question is: "How much functionality should we pack in our smart pixels?" The author will not attempt to answer this question, because the answer is a function of many variables related to device packaging. In actuality, the question of optics vs. electronics really boils down to a comparison of different device packaging technologies. Device packages for computing systems must provide three fundamental features: 1) stable mechanical mounts for the devices, 2) adequate thermal paths for heat removal, and 3) adequate signal paths between the devices. The choice of photons or electrons for the last feature is dependent on several parameters including the bandwidth of the signals, the number of signals that must be transmitted between the devices, the distance between the devices, and the overall system architecture. The variability in system architecture complicates the problem the most, because different architectures can be proposed to take advantage of the different capabilties of the different technologies, so we often have to compare apples and oranges to determine the best solution. Each system must be analyzed separately, and the answer to the above question can be found only after the feasibility of electronic packaging solutions have been investigated. Improvements in electronic packaging (such as pin grid arrays, flip-chip mounting, and multi-chip modules on ceramic of silicon hybrids) have greatly increased the capabilities of electronic packaging, so optical solutions must be even better if they are going to be cost effective alternatives to the electronic approaches.

Researchers are still hunting for the holy grail of architectures in free-space digital optical computing, and it has been very elusive. Although many interesting architectures have been developed, the author does not believe that any of them truly capitalize on all of the potential benefits of optics (bandwidth and parallelism) in an efficient, cost-effective manner. The holy grail of architectures may be out there, but it may be so different from anything that we are familiar with that it may take another genius like Von Neumann to discover it. Nevertheless, some of the architectures that have been borrowed from the electronic world do seem to capitalize on some of the benefits in optics. In particular, neural networks, which require large degrees of connectivity between small neural processors, seem to be a very promising candidate. In addition, computing architectures that require large, high-bandwidth switching networks might also be worth consideration in future research efforts, because these systems can draw on related work in photonic switching. For example, data flow machines, which require large degrees of connectivity in a switching network between firing memories and processors, could conceivably use photonic switching. In addition, the entire field of distributed processing may be able to take advantage of the capabilities of photonic switching to provide high-density, high-bandwidth connections between processing elements. This capability could permit an entire memory in one processing element to be transmitted in parallel via free-space optics to a memory in another processing element. Image processing is a typical example of an application that might require this capability.

In conclusion, the promise of optics is real, but the work required to captitalize on that promise is still in its infancy. In the opinion of the author, only a coordinated effort between system architects, device and laser researchers, optical engineers, and opto-mechanical designers will yield a useful result. In addition, the works must also be coordinated with the high-speed electronic designers, because hybrid optical/electronic systems with optical interconnections between small localized electronic processors will probably yield the largest pay-back.

Spectral Processing

Alan Craig
Air Force Office of Scientific Research
Building 410 Bolling AFB
Washington D.C., 20332-6448
craig@ccf2.nrl.navy.mil

An electromagnetic wave in the near infrared region of the optical spectrum has a frequency of 200-300 terahertz (THz). These waves are often used as carriers for information transmission of signals whose modulation bandwidth seldom exceeds 1 gigahertz (GHz). This modulation bandwidth is a miniscule fraction of the carrier frequency. In many modulated carrier systems (*e.g.* radio or television) the modulation bandwidth exceeds 25% of the carrier frequency. To overcome this deficiency, telecommunications researchers have turned to frequency division multiplexing (FDM) and its wider-carrier-separation counterpart wavelength division multiplexing (WDM). In these systems, each frequency or wavelength designates a definite single channel between a signal's source and its receiver. Information is impressed on this channel by temporally modulating the optical carrier at the designated wavelength at the highest rate possible with currently available technology (subject to cost constraints); often, several signals with identical source-destination terminals are time-division-multiplexed on a single carrier to use all of the available modulation bandwidth (up to several gigahertz).

For computer interconnects, this configuration of the broad, accessible, wavelength bandwidth may be less than optimal. Computer busses and backplanes are configured to transmit data in byte-at-a-time, bit-parallel formats on ribbon interconnects 16 or 32 wires (or fibers) wide. Interfacing to the bit-serial format of FDM or WDM transmission requires buffering and sequential to parallel (or vice versa) conversion at each transmitter and receiver node.

An alternative transmission format that makes use of the identical carrier capacity, but allocated differently, proves a better match to computer interconnect bit-parallel formats. In this approach, several neighboring wavelength channels, *e.g.* 16, are assigned to carry a single binary word, with each wavelength representing a binary phase. Rather than sending 16 temporally encoded signals, each on a single channel, modulated at 1 Gbit/sec rates, a single byte of 1ns duration occupies all 16 channels simultaneously. This wavelength-encoded byte arrives simultaneously, *i.e.* in parallel. This matches processor computation design and capability.

A straightforward engineering approach to wavelength encoded transmission makes use of many of the same components as WDM, controlled by different algorithms: tunable, or selectable-wavelength arrays of semiconductor lasers; wavelength sensitive switches; tuned or tunable wavelength filters. (Admittedly, making these components conform to control by the new algorithms and providing the appropriate control signals may be challenging.) To look beyond this capability, consider the prospect of not only transmitting in this wavelength encoded format but also performing decision processes optically or opto-electronically in this realm. Specifically, this pursuit implies a system in which information is carried not according to amplitude levels, but according to wavelength selection. Processing then entails controlling the output wavelength(s) of a node by some device operation that is dependent on the input wavelength(s) to the node.

Amplitude may be required to cause a process to occur, but the control and the input and output data are all wavelength encoded.

New devices will be required to do this processing. Several of these have been envisioned, and research programs are initiated to determine their viability. Laser diodes whose emission wavelength can be controlled by illuminations with or injection of light at a difficult wavelength are conceived. Vertical cavity arrays of laser diodes in which the emission wavelength of the lasers is raster-stepped with uniform separation across the array have been built; addressing remains to be resolved. Wavelength sensitive coupled-mode and interferometric switches are being developed. Wavelength (or frequency) conversion nonlinear optic effects in resonators and waveguides to perform 3-wave and 4-wave mixing and optical parametric processes in organic, crystalline and semiconductor materials are being investigated. Tunable filters may result from laser-diode amplifiers or from spectral-hole-burning absorption filters in semiconductor quantum box materials.

These devices, or simple combinations of them, provide various transfer functions where optical wavelength both indicates the input and output states and provides the operation control. Some capabilities of these wavelength-encoded processors can be seen immediately and can provide features difficult to realize in present day amplitude modulation devices. Compare and conditional branching, or table lookup may be realizable in these devices. A list of conceivably implementable, computationally useful primitive functions follows:

- Boolean functions
- Compare, shift, invert, perform 2's complement function
- Algebraic functions
- Transcendental and trigonometric functions
- Powers and logarithms
- Linear algebraic functions
- Table lookup and other database operations
- Switching and routing
- Operations of medium complexity (*e.g.* combinations of Boolean functions)

This list is common to all computing architectures. A study is under way to determine which if any might be performed efficiently, accurately, and quickly in the wavelength-encoded format by these new devices.

This is an exploratory research area whose efficacy is not proven. Skepticism about its prospects is healthy, but enthusiasm is welcome. If nothing further, it is hoped that these concepts and attempts at building and controlling new devices will contribute to realizations about computing and data transmission algorithms and to capabilities in device physics heretofore unprobed.

Why Optics have a Part to Play in Von Neumann Machines; and why Von Neumann Machines have no Part to Play in Optics

Alex Dickinson
AT&T Bell Labs

It has been long recognized that system speeds are not so much limited by the speed of devices, but more the speed of the interconnect between devices. As line widths and hence device sizes have scaled down, average interconnect lengths have remained more or less constant, increasing the effective cost of on-chip interconnect. Worse still, as the speed and number of devices on a chip have increased, the demands on chip i/o have increased correspondingly (Rent's rule implying increasing number of i/o; higher clock rates necessitating increased i/o speeds).

A Von Neumann architecture is particularly vulnerable to interchip connection bottlenecks. As instructions are executed faster, greater bandwidths are required of the processor-memory link. Additional caching on the CPU chip helps to a point, but as code sizes continue to increase in proportion to processor speed, it is difficult to contain increasingly large portions of the code in cache without it occupying large chunks of valuable CPU chip real estate.

Quantum optical devices and manufacturable optical systems may offer a solution to this problem. A quantum modulator such as a SEED, or a low threshold laser such as the SEL, allows us to generate modulated beams in a very small space with a very small amount of power. Suitable detectors and amplifier circuits allow us to decode the signals on another chip after the beam has been transmitted through some free space optical system.

We envision a machine in which high speed links of many channels in a regular topology (such as the processor-memory bus) are carried optically within a planar glass substate. Other signal and power runners are fabricated from metal layers deposited on the substrate, and chips are bonded in turn onto these runners. We are presently carrying out fabrication experiments and architectural simulations for this class of system.

This form of construction offers one of the most direct paths for optics to make a contribution to Von Neumann computer practice. But is there a converse path then along which Von Neumann computer structures can make a contribution to the much sort after optical computer?

It would seem unlikely.

The primary advantage of optics over electronics appears to be in the ability to create large arrays of fast, regular, and efficient interconnect. Now, although Von Neumann machines exhibit the need for such connections in certain places (such as that described above) they do not in general exhibit a great deal of regularity at the gate-to-gate level. Attempts to map Von Neumann machines into this regular regime require that many connections and devices be abandoned with few if any compensating advantages.

We are developing a non-Von Neumann architecture comprised of a large number of identical, simple, finite state machines. Each FSM executes combinator code in parallel, and is amenable to implementation with symbolic substitution. This structure should be a much more appropriate match to the digital optical technolgy that is becoming available by virtue of it's low level simplicity and regularity.

We are intending to construct a VLSI version of the architecture in the summer of 1991 as a means of gaining more understanding of the issues that would be involved in an optical implementation.

Timothy J. Drabik
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Optoelectronic Processor Arrays

Free-space optics should be used to interconnect wafer-scale integrated (WSI) or ultra-large-scale integrated (ULSI) parallel processor arrays. Two long-standing problems can be addressed with an optical approach: (1) All but locally connected topologies suffer from the poor performance of cross-wafer electrical interconnects. Inter-processing-element (PE) optical interconnects could give cross-wafer paths satisfactory performance, as well as reduce the silicon area required to implement a given array size for small-diameter, high-wire-area topologies. (2) Providing reconfigurability and redundancy for defect tolerance has made it difficult for WSI to utilize more than 50% of the wafer area. Sparing strategies that rely on the availability of extra PEs within a local neighborhood of a defective PE can be overloaded by clusters of defects. If reconfigurability is provided by an optical interconnect system, sparing can be less constrained, and apart from the need to provide enough functional PEs, no reconfigurability need be designed into the circuitry itself.

Diffraction-based analysis of the capacity of optical interconnects yields a reciprocal relationship between acuity and interconnect complexity: an optical system occupying a given volume can be used to interconnect many closely spaced PEs in a simple way (imaging, for example) or fewer, larger PEs in a more complex (more space-variant) way. PEs with diameters of the order of 1 mm, and having one light modulator each, represent a sufficiently low requirement on imaging acuity as to admit consideration of very complex topologies, with capacity left over for reconfiguration around defects.

Optically interconnected WSI can be considered as a means to approach the asymptotic volume density of 3-D VLSI using only planar (e.g., 2-level metal) technology. As an example, consider $n \log_2 n$ PEs connected in the butterfly topology, and allow n to grow. In a fabrication technology with a fixed number of interconnect layers, Ullman has shown that $\Omega(n^2)$ area is required. In an hypothetical full 3-D (isotropic) technology, where elements can extend in the vertical direction as easily as in the other two, Leighton and Rosenberg have shown that the volume of a butterfly grows as $\Omega(n^{3/2} \log^{1/2} n)$. Analysis has shown that an optically interconnected system using circuit technology with two levels of metal requires only $O(n^{3/2} \log^2 n)$ volume.

"All-optical" free-space architectures are subject to serious, fundamental physical constraints on performance. Diffraction bounds the area density of computing elements, and minimum feedback latency is determined by the optical path length. By contrast, a VLSI-based PE array might be constrained by the imaging configuration to have its light modulators 1 mm apart; however the number of gates in the 1 mm^2 PE is free to track advances in fabrication technology. Moreover, while the communication latency in an optically interconnected VLSI-based processor is certainly established by optical path length, tight feedback loops (such as the carry operation in a bit-serial adder) can be pulled into an electronic PE and thus be free of this constraint. Falling into the all-optical category are thus not only those systems with intrinsically optical computing elements, but also those using electrically based elements with small, fixed computational capability (like SEED devices), because the number of gates per light modulator is fixed at approximately 1.

Opto-Electronic Versus All Optical Computers

Sadik Esener
Ece Department, R-007
University of California, San Diego
La Jolla, Ca 92093
(619) 534-2732

Over the past few years, optical interconnection technology has penetrated the computer industry. Several computers that use optical interconnections have been announced, or are already on the market. This indicates that we are quickly entering the era of hybrid opto-electronic computers. In such computers, signals propagating between switches are carried out both by electrons in conducting wires as well as by photons that do not require conductors. By combining the complementary strengths of both optical and electronic technologies, hybrid opto-electronic computers provide optimal solutions for the implementation of increasingly parallel computers. Presently, it is of great interest to examine how computer architectures and performance will be affected by the increasing use of optics. This is strongly dependent on the level of computer packaging at which optical interconnects are used. A natural question that arises is whether an all optical digital computer will ever become useful?

As can be seen from Fig.1, hybrid opto-electronic computers encompass on one extreme all electronic computers and on the other extreme all optical computers. In the latter, communication between switches are carried out by photons alone. Such computers can be considered as special cases of hybrid opto-electronic computers. For the design and optimization of any opto-electronic computer it is important to establish a quantitative measure of the respective amounts of computation and communication work carried out by the optical and electronic components. To this end, at UCSD, we utilize the concept of the grain size of an OE processing element (PE) which is strongly related to the ratio of the number of electronic gates in the system to the number of optical light transmitters. For best performance the grain size is optimized with respect to bandwidth, system size and power dissipation. Thus the grain size and therefore the relative role of optics in computers is strongly dependent on the specific application and available device technology.

Since the use of optical interconnections in computers enables better implementation of global and dense communication, it is natural to expect that the implementation of multistage interconnection networks (MINs) will require the highest usage of optical techniques. The grain size study on the opto-electronic implementation of MINs using free-space optical interconnections that was carried out at UCSD clearly demonstrated that this application together with today's state of the art device characteristics would lead to an optimal grain size where for every light transmitters there will be roughly 200 corresponding electronic logic gates.

For such a small grain size one may ask whether an all optical approach would not be better suited. The benefits of such an approach would be the utilization of a lower cost device technology without requiring parallel accessed opto-electronic integrated circuits (OEICs) where detectors, logic circuits and light transmitters must be integrated. When the asymptotic behavior of the graph in Fig.2 is analyzed, one concludes that under the

performance requirements and assumptions made in Ref.1 the power dissipation of an all optical implementation would be prohibitively large. This power dissipation indicates that the switching energies of our state-of-the-art optical devices are not sufficiently small to make an all optical approach viable.

Is an all optical computer ever going to be viable? To become viable in short term with existing devices, novel applications that further reduce the optimal grain size are required. These may well be in the area of long distance communication switching. Such applications must be uncovered and analyzed before attempts are made to build systems. In the long term, with significant progress in optical switching devices in terms of speed (exceeding 10GHz), array size (exceeding one million), lower switching power and improved power removal, one may consider this approach as a potential alternative to hybrid opto-electronic computers. However, it should not be forgotten that opto-electronic computers will continually leverage on the progress made both in electronic and photonic domains and will be very competitive once a low cost technology base is fully developed.

Reference

1. F. Kiamilev, A. Krishnamoorthy, P Marchand, S. Esener and S.H.Lee "Design of interconnection networks for programmable optoelectronic multiprocessors" Proc.of Topical meeting on Optical Computing of OSA, Kobe, April 1990.

Taxonomy of Opto-electronic Computers

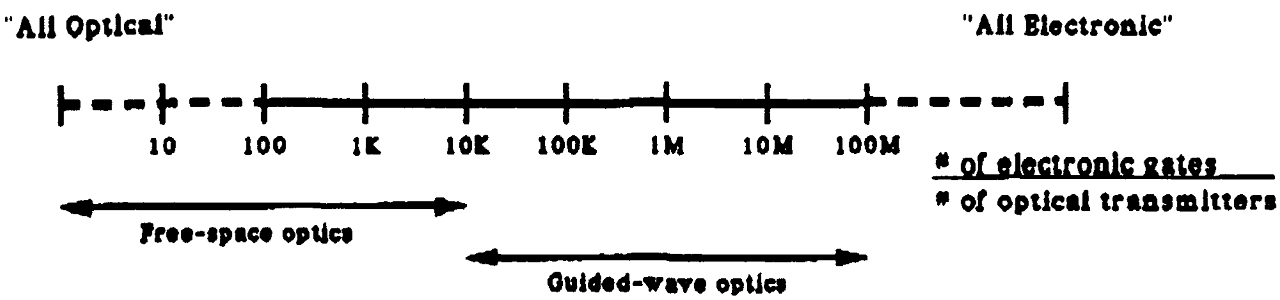


Figure 1

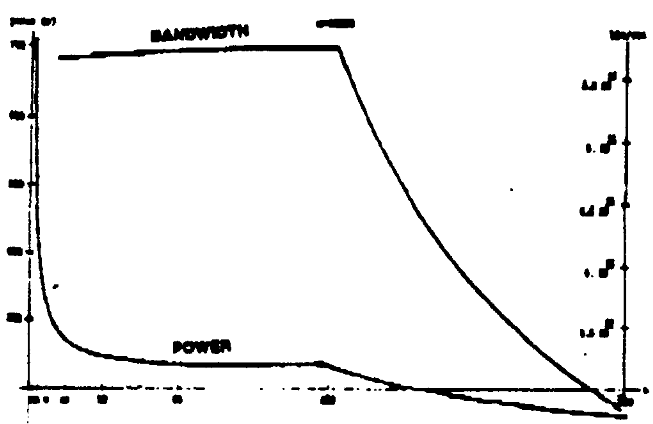


Figure 2a

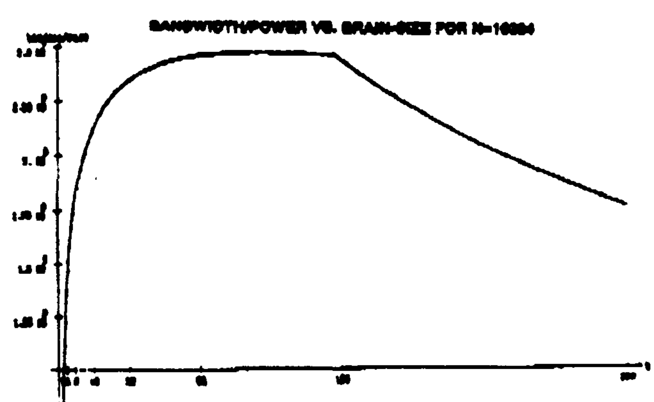


Figure 2b

A Position Paper on Future Directions in Optical Computing

C. Lee Giles
NEC Research Institute

January, 1991

We present a perspective on future directions in optical computing by giving a critical interpretation of the past successes and failures in optical computing and processing and then speculate on fruitful near and long term research directions. The danger of making such predictions is well-understood and the reader is justly cautioned. Furthermore, these arguments are solely the personal opinion of the author and in no way reflect those of NEC. We interpret the goal of optical computing to be to eventually develop optical devices and/or systems that will be actively used in real computers or computing problems. Therefore, our definition of an optical computing success is the actual use of optical devices/systems in "many" real computers or computing problems (this excludes one-of-a-kind successes). A distinction is made between near-term goals, those that will fit into actual computing or processing in the next 5 to 10 years, and long-term goals, the exploration of new computing models, a current example would be neural networks. [Because a research direction is defined as long-term does not imply that research should not be currently encouraged or supported.]

Near-term goals: Our view is based on the assumption that current computing technologies (silicon, III-V's, etc.) and architectures (von Neumann, multiprocessors, etc.) are far from obsolete and will continue to dominate computation well into the next century and most likely beyond, and that the most profitable near-term approach for optics is to successfully "fit in" to these technologies. [For example, the microprocessor industry conservatively predicts that using reasonable extensions of existing technologies, the microprocessor in the year 2000 will be a 64 bit, 1/4 GHz, GIP machine with a 6 square cm active area and 0.1 micron line resolution. These projections totally ignore the most recent progress in memory-based wafer scale integration and significant improvements in chip power dissipation.] We contend that existing technologies are very difficult to replace unless the replacing technology offers not only improved performance characteristics, but also such mundane aspects as improved cost, ruggedness, manufacturability, integrability, etc. In replacing an existing technology, one can choose to replace all or part(s) of it. When a technology is just about peaked in performance, then replacing it with alternate technologies seems reasonable (ex. steam engines replacing animals as sources of power.). For an unpeaked technology - silicon, III-V's - it seems more reasonable to look for methods which now or in the future will drastically improve total system performance. This seems to be why optical interconnects and memory have been successes, they readily fit into existing technologies and offer improved performance. But why have optical signal processors and pattern recognizers not been successful? Memory and interconnects are basic hardware primitives innate to all computation; Fourier-based processing and pattern recognition are specialized complex operations. We argue that a replacement technology stands a better chance of success if it replaces primitive, widely-used operations instead of complex specialized ones. Of course today's complex specialized operations could be commonplace tomorrow. At present signal processing is either done in software and in specialized DSP chips that are programmable and capable of performing a multitude of signal processing operations - not the few specialized operations of optical signal processing systems. What is not of issue here is also important -whether a process is analog is of no importance as long as it integrates into existing systems.

To summarize, we argue that optics will have the best opportunities at short term successes if the optical process offers enhanced performance to existing or projected systems. For example, with the continuing integration of traditional technologies due to increasing size and density, optical interconnections with its enormous bandwidths seems a certain winner.

Long-term goals: The future directions of computing seem to be best described by the descriptions - "more" and "friendly." We will discuss only the "more" - i.e. more speed, more memory, more power per computer, more computers (parallelism, usually defined as a multitude of computers working together in some productive fashion). We interpret the "friendly" part of computation to be a function of software and not hardware, of course realizing that the software might be highly dependent on the power of the hardware available. We feel that optics has little if any role to play in software. Many computer scientists predict that for the next decade and beyond, parallel computation will start to have a significant influence on computation, regardless of existing machines and their long lifespans. It would seem that optics should look for roles it can play in parallel computation - from hardware inroads to a potential impact on parallel computer architecture design. This is a direction that is currently being pursued to a limited extent in optical computing. However, an essential question relates to the old cart and horse problem. Can architecture considerations have much impact and meaning when the hardware doesn't exist? Probably not. Certainly some optical computing architecture research is motivating, but it will soon end without existing hardware. Again, we make a similar argument that the successes of optics in parallel computation will come from optical processes enhancing performance of parallel computers. Designing a parallel computer is an extensive "group" task that not only includes hardware but software considerations. Unless the *optical computing* community has significant hardware successes, it would be futile to undertake a "separate" optical parallel computer design. The alternative is to use the leverage of existing parallel computers and to coordinate research and design with these programs. Thus, we contend that long-term optical computing should focus on integrating its research into the traditional directions of parallel computing. This integration could take on many directions, from data flow machines to neural networks.

Conclusions: We contend that the field of optical computing would best be served by focusing primarily on optical processes that enhance the performance of existing, planned and future digital computers, whether they be serial or parallel. These optical processes, either digital or analog, should be appropriate and crucial for the computer architecture for which they are planned. The optical subsystems or components would stand the best chance of use and success if they are as computationally primitive as possible [this is certainly a function of the type of computer and what is primitive now might not be so in the future]. We also believe that the field is well served if some selected research is directed toward new directions in computation or computing, so that the further use of emerging optical computing technologies is continuously encouraged.

C. Lee Giles
NEC Research Institute
4 Independence Way
Princeton, NJ 08540
USA

Internet: giles@research.nj.nec.com
UUCP: princeton!nec!giles
Phone: (609) 951-2642
FAX: (609) 951-2482

Position Paper: Let's Not Neglect the Architectures

Mike Haney
BDM International, Inc.

There has been a great deal of recent excitement in Optical Computing (OC) about the rapid advances made in component technology and the significant engineering accomplishments being made in demonstrations. Furthermore, there has evolved a new and interesting emphasis on the application of CAD tools to the design of optical computers. These developments are extremely important to the eventual success of OC. However, I fear that there has been a concurrent decrease in attention to the development of architectures, and, more specifically, the mapping of computing problems onto them. This inattention seems to have occurred at exactly the wrong time -- before a clear, quantitative case is made for OC vis-a-vis specific computing problems. Even more troubling is the attitude, expressed by some, that architectural developments will follow "easily," once the technology has arrived.

There are several reasons for this situation, not the least of which are the many years of "paper architectures" that were often not well thought out and not mapped onto problems to show their significance. There developed a real (and somewhat justified) disparagement toward the continuous stream of new architectures. Consequently, with the pace of device technology rapidly accelerating, a "Just do it!" attitude developed -- almost as a knee-jerk reaction and despite the lack of a clearly or even vaguely defined path to ultimate success. The emphasis in OC has therefore rapidly swung from the domain of the extremely theoretical and broadly focused computer architects to that of the experimental and narrowly focused physicists and engineers.

Neither of these situations is healthy, but the new developments might harbor more danger to the OC field. In the

former situation we were subject to the criticism: "Yes, but what does it mean?" Now we are in a more precarious position. If we are unable to make convincing arguments about the extensibility of the architectures we demonstrate, then the criticism might become: "Is that all there is?"

There needs to be renewed attention to the middle ground. We must concentrate on developing architectural concepts on which real problems can be mapped. We need applied computer architects that measure success by a computer's performance on well-defined problems or benchmarks rather than by general metrics such as throughput and number of interconnections, whose true meaning are difficult to discern. To accomplish this, the OC field needs more generalists -- to link the new component technologies together to solve specific problems. Perhaps new CAD tools can help make the arguments, but some first order analysis should also be possible without them. The applied computer architects must strive to divorce themselves from favorite technologies or architectural notions.

In summary, we need to adopt a "focused top-down" perspective in OC. By this I mean that we should begin with specific computing problems (not broad problem types) on which the performance of electronic computers are already well characterized. The advantages of proposed photonic solutions can then be evaluated more readily. We should strive to make statements like the following: "If technology X is applied in architecture Y, then problem Z can be solved with projected performance enhancements A,B, and C."

Can Optical Switches Make the World a Better Place to Compute In?

Saul Levy
Department of Computer Science
Rutgers University

This meeting is addressing problems of using digital optical switches. From what I've heard, the assumption seems to have been made that as soon as we produce a good digital optical gate, the world of computer architects will beat a path to our door. Everyone here knows the promises of optics: great communication, enormous parallelism, amazing switching speeds, even more remarkable storage densities. How does this promise compare to the promise of existing technologies? We must never underestimate the built-in advantages of an existing technology; there are large numbers of specialists (some of whom are very bright) who understand the technology, there is a major effort to improve the technology incrementally (often in very large increments), and probably most important of all there is a large infrastructure of very smart and committed people who provide all the ancillary hardware and software which support the current technologies. The technology of television and video terminals is very close; they both use the same video tubes, and much of the same analog display circuitry, and there are tens of millions of television sets in this country. When a digital high definition television system is adopted, it will surely be all electronic, and might even induce a serious return to the manufacture of memory chips on the part of domestic ic manufacturers. The enormous volume of digital electronic circuits that will be generated in response to that new technology will drive costs for those components down very rapidly. And even if the new HDTV is slow in coming, the current technology is improving at an absolutely amazing rate, and has sustained that rate of improvement over a very long period. Growth of the performance of computers has increased at an annual rate of 20% to 35% per year for the last 20 years, semiconductor memory densities have been quadrupling every three years, with prices of DRAM chips dropping at 40% per year (till they level off at about \$1 independent of size)[1]. Even when one considers disk

[1] Hennessy, John L. & David A Patterson *Computer Architecture A Quantitative Approach*. Morgan Kauffman ,1990

storage (where optical devices have an obvious physical advantage and a thriving infrastructure), magnetic disks have achieved storage densities as great as optical disks, and read/write/access characteristics which are far superior to their optical counterparts, and they still continue to improve at a rapid rate.

There is a great deal of work on architectures which use the current technologies. System architects are comfortable with the characteristics of the technology, and have a sufficient understanding of the limitations of the technology to design around them. Barring a miracle (or some other substantial occurrence) the current architects will determine what the next generation of computers will look like. I think it's clear, even to the most optimistic among us, that, if we hope to impact computer architectures, we have to first find a niche, a place where the current technology is weak. Then, using that niche as an opening wedge, we can introduce optical technology to the designers as a *solution* to a problem they really *believe* exists. Right now, there is a tremendous air of confidence that improvements in the current technologies will lead to teraops computing in just a few years, with the use of clever *massively parallel* architectures. Fortunately for us, a critical physical problem in constructing massively parallel computers is massively parallel communication, a problem for which optics is uniquely well suited. So I believe that's our opening wedge. We can hook them on optical interconnections. But how will we follow up? What else can we do better with optics than they can do with electronics? We talk about faster switching, but their switches are getting faster every day, and to switch fast we'll need large amounts of optical power (which is difficult to come by, and may require low duty cycles); so it's not clear we have a winner there. But what we do have is the precise control of the movement of data between switches. Electronic circuits can only be reliably pipelined in relatively complex chunks because of the need to insert buffers between pipeline stages which increase the latency of the operation, limiting the potential speedup due to use of the pipeline. We, on the other hand, can control the data flow so precisely that we can consider pipelining at the individual gate or switch level without requiring any buffers between the pipeline stages.

Spatially Multiplexed Optical Content-Addressable-Memory

For Digital Optical Array Logic and Arithmetic Processing

Yao Li

Department of Electrical Engineering,

City College of New York, New York, NY 10031.

I. Background

The use of a programmable array logic concept in optical computing where optics is mainly used for switching and routing rather than a direct logic gating has recently been recognized [1-7]. An N variable logic function can be expressed as

$$f(x_1, x_2, \cdots x_N) = \sum_{i=1}^{k} [\hat{x}_1 \bullet \hat{x}_2 \bullet \cdots \hat{x}_N]_i \tag{1}$$

where $1 \leq k \leq 2^{N-1}$, $\sum$, $\bullet$ and $\hat{}$ denote an OR, an AND, and one of the two states of a variable, respectively. Eq (1) which implies a logic sum of various N-variable logic product terms can be further decomposed, using a DeMorgan theorem, to expressions containing smaller sized logic product terms. Depending on the available optical hardware, the selection of the product term size for optical programmable logic implementations also varies. The minimum size (two variable) product terms have been used in shadow-casting-based, and symbolic-substitution-based array logic approaches. One problem of using a small-sized product term is that to process a N-variable function many cascading stages are needed which in turn slows down the overall processing speed and introduces cascading related problems.

II. Optical Content Addressable Memory (CAM) Array Logic

To overcome the processing slow-down and "over-cascading" caused by the use of small product term sizes, the optical CAM approach we are currently investigating uses large (limited by the hardware dynamic range) product terms directly. As an example of generating an 8-variable product term $(x_1 \bullet x_2 \bullet \bar{x}_3 \bullet\ d_4 \bullet \bar{x}_5 \bullet x_6 \bullet d_7 \bullet d_8)$, a schematic encoding and computing process is shown in Fig.1.

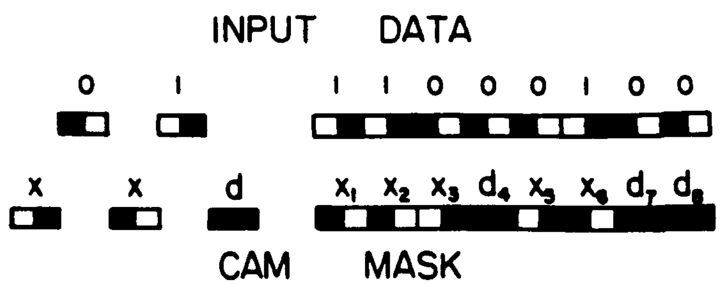


Fig.1. Input and CAM encodings.

Here, d denotes a "don't care" which can be either "1" or "0". To generate an output 1 from this product term, eight input possibilities which are 11000100, 11000101, 11000110, 11000111, 11010100, 11010101, 11010110, and 11010111. A standard dual-rail input encoding is shown in the left-hand side. A typical input encoded data (the first of the eight given expressions) is shown in the top right-hand side. The coded CAM mask designed to incorporate all eight input possibilities is shown in the bottom of the right-hand side. When the input pattern carrying one of the eight searched inputs illuminates the CAM mask, a "0" light is detected which can be electronically thresholded and inverted to generate a logic "1" output. Inputs other than the eight given will generate residue light at the detector which, after a threshold and an inversion, corresponds to a "0". The term CAM [3] is used not only because

of the approach reduces the logic information to its most compact form, but also because the function's output contents, e.g. "1" and "0", rather than its input locations are used for addressing. The accuracy of this method for processing a large N product term depends on the dynamic range of the analog optical and electronic components employed. Fast electronic threshold detectors with large dynamic range performs a crucial role in this approach.

Now, to process k product terms each having N variables, an optical vector-matrix processor architecture [4,7] can be employed [see Fig.2]. While the matrix represents an array of k 1D coded CAM mask sequences, the input vector serves as the common input to all k CAM masks. A 1D optical analog intensity summation at the output generates k matching results for the electronic threshold detector. The major difference between the CAM and conventional optical analog matrix processing is the result treatment at the output. With the optical matrix algebra where the output needs to be A/D converted, a low accuracy is inevitable, while with the optical CAM where only a "0" light needs to be distinguished from the other values, a higher processing accuracy can be expected.

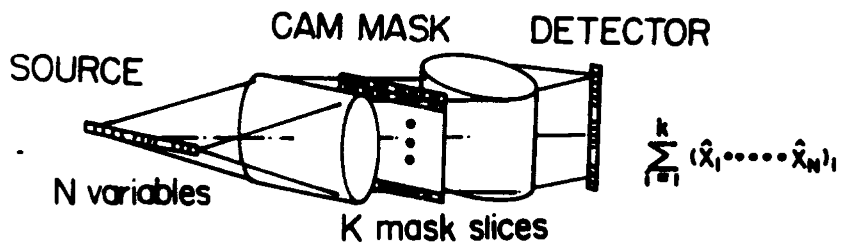


Fig.2. Optical vector-matrix multiplication for CAM processing.

III. Spatially Multiplexed Optical CAM Processing

The difference between an optical and an electronic CAM implemented through a programmable logic array is that the former uses 3D free-space optics to replace the 2D wiring pattern of the latter. The advantage thus gained by using optics may not be convincing enough for adopting an optical approach. However, when the optical CAM is used together with a spatial multiplexing scheme, a feature that electronics has no way to realize, some obvious advantages appear. By spatial multiplexing we mean that the 2D CAM mask is shared among a group input vectors so that when different data are processed for the same application, no proportional amount of space extension is required. In a SIMD environment,

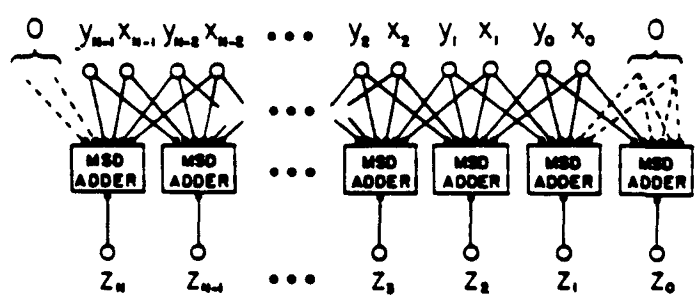


Fig.3. A flow diagram for a N-bit parallel MSD adder.

multiple data are identically processed upon execution of an identical instruction. A typical numerical example is to add two N bit MSD numbers where MSD adder units acquires six

inputs to generate each bit addition result (see Fig.3). The use of free-space optical CAM allows to generate the parallel MSD addition result on-the-fly by integrating many vector-matrix product processor to one matrix-matrix product system. On the other hand, when the identical processing task is handled by an electronic CAM approach, the repetitive use of hardware is inevitable. The recent breakthrough in the multiple matrix multiplication schemes has provided a technological base to implement a spatially multiplexed CAM logic and arithmetic processor [8-9]. In Fig.4, one typical approach for a fully parallel N channel multiplexed CAM-based MSD adder is depicted. The required matrix-matrix multiplication is performed through a triple matrix product processor by setting one of the three matrices to an unity matrix.

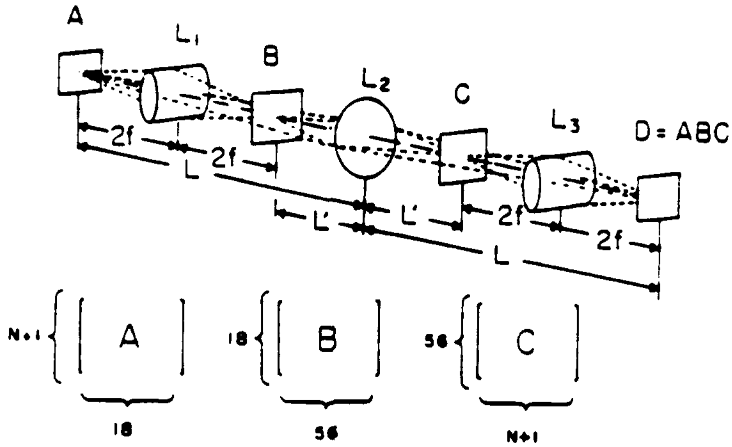


Fig.4. Spatially multiplexed (through matrix-matrix multiplication) optical CAM MSD adder.

IV. Conclusions

It is the size of the optical logic product term that determines the processing speed and efficiency. Using the available optical and electronic hardware, it is preferred to implement optical programmable logic arrays with large logic product terms. For an optical programmable logic array to be efficiently used, it should be designed in such a way that can process "don't care" variables. The logic contents rather than locations should be used for addressing. To incorporate unique advantages of optical processing, spatial multiplexing should be considered in the design. The spatially multiplexed optical CAM processor will find many applications in SIMD processing where different input data needs to be identically processed.

V. Acknowledgement

This work is supported in part by the Air Force Office of Scientific Research and by the National Science Foundation.

VI. References

[1] A. A. Sawchuk and T. C. Strand, Proc. IEEE 72, 758 (1984).

[2] Y. Ichioka and J. Tanida, Proc. IEEE 72, 187 (1984).

[3] M. M. Mirsalehi and T. K. Gaylord, Appl. Opt. 25, 3078 (1986).

[4] P. S. Guilfoyle and W. I. Wiley, Appl. Opt. 27, 1661 (1987).

[5] Y. Li, A. Kostrzewski, D. H. Kim, and G. Eichmann, Opt. Lett. 13, 895 (1988).

[6] M. J. Murdocca, Ph.D. thesis (State University of New Jersey, 1988).

[7] Y. Li, D. H. Kim, A. Kostrzewski, and G. Eichmann, Opt. Lett. 14, 1254 (1989).

[8] H. Nakano and K. Hotate, Appl. Opt. 26, 917 (1987).

[9] H. Huang, L. Liu, and Z. Wang, Opt. Lett. 15, 1085 (1990).

A New Optical Interconnection Technique For Architectures Involving Wafer-Scale Arrays

Richard A. Linke
NEC Research Institute
4 Independence Way Princeton, New Jersey 08540

Introduction

It has been argued that optical techniques can provide interconnection networks for complex electronic systems with important advantages over present electronic interconnection techniques [e.g., ref 1]. These advantages include increased interconnection density and bandwidth, as well as the potential for dynamic reconfigurability. One property of optics which provides some of the justification for such optimism is the non-interacting nature of light: unlike electrical signals in wires, optical signals can pass through each other without interfering.

Unfortunately, this non-interaction feature of light is not exploited in most waveguide-based networks which use optical fiber or integrated optical waveguides and switches. In these architectures, individual signals are confined to unique waveguides. The property is used to advantage in some free-space interconnection schemes which employ lenses or holographic imaging techniques. These architectures, however, share another problem: propagation path lengths are necessarily comparable to the array linear dimensions to avoid impractically large numerical apertures. For the case of interconnections between planar, N X N, arrays of elements, for example, this means that not only will propagation delays increase with N, but also the packing efficiency (elements per unit volume in a multi-stage system) will fall as 1/N.

A Planar Broadcast Network

An architecture which takes advantage of the non-interacting nature of light and, at the same time, allows for very high packing densities, is shown in Figure 1. The figure depicts a semiconductor-wafer scale network based on the use of a two-dimensional planar optical waveguide as a broadcast medium [2]. The proposed network has many of the advantages of optical broadcast networks based on a star coupler [3] and, in fact, an integrated optical waveguide star coupler network could be used in place of the planar waveguide, though at considerable increase in complexity and size. Anticipated disadvantages of the planar guide, compared with the complete star coupler network, are the relatively inefficient use of optical energy and a wide dynamic-range requirement for the optical receivers.

Figure 1(a) depicts the surface of a semiconductor wafer containing a planar array of electronic processing elements (PE's) each of which is in electrical contact with an optical source and an optical detector used for communications with other PE's (as well as with specialized input and output elements which are not shown). Assuming, for example, a 4-inch diameter semiconductor wafer divided into 256 local computation regions gives each region an area of approximately 25 square millimeters, or about that of a current reasonably powerful microprocessor chip. This area will contain electronic logic and storage as well as an optical transmitter and receiver. The electrical drive power requirements of the optical source and electronics will be approximately equal.

Figure 1(b) shows the various layers including the optical guiding layer in which light propagates in much the same way that radio waves propagate along the surface of

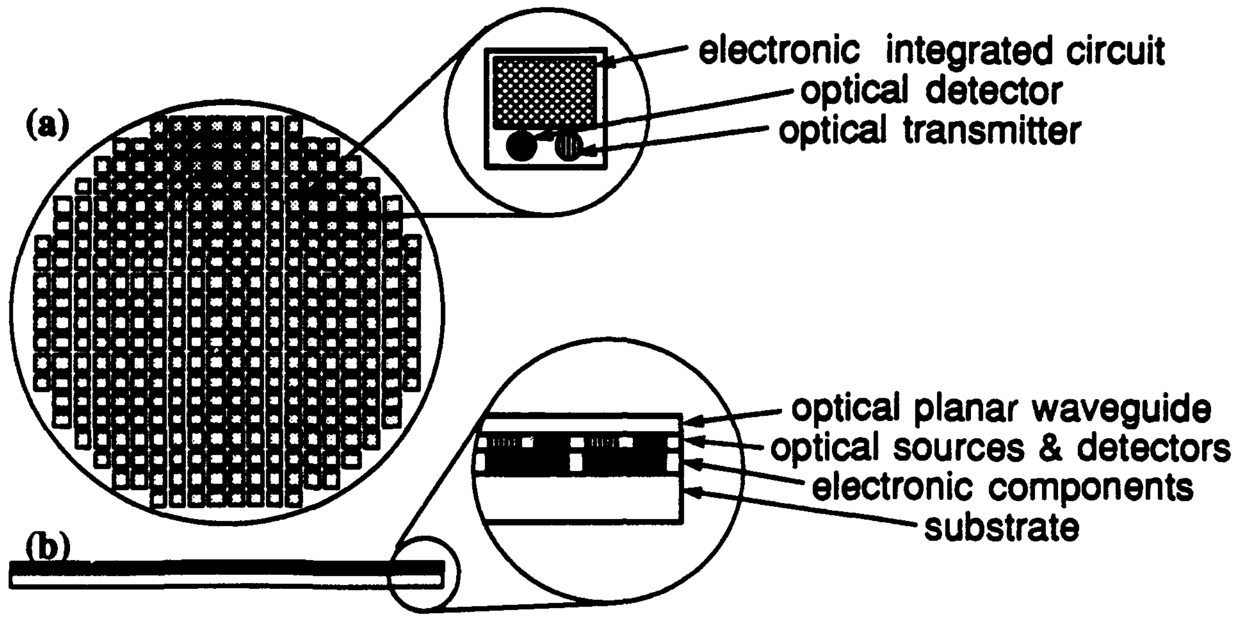


Figure 1. *Top(a) and edge(b) views of a proposed wafer-scale network of electronic circuits (e.g. microprocessors) interconnected via a two-dimensional optical waveguide. The planar waveguide distributes signals from each optical transmitter to all receivers in a broadcast and select mode.*

the earth: light coupled omnidirectionally into the guide at any point on the disc is broadcast to every element on the wafer. The detected signal strength will fall as 1/d, where d is the separation between source and detector. For a 1 millimeter diameter detector, the maximum loss due to signal spreading is about 25 dB in this example. Assuming an additional 5 dB for propagation and coupling losses and a coupled transmitter power of 1 mW, this means that the receiver will detect at least 1 µW (-30 dBm).

The resulting network might be operated in a manner similar to that of an Ethernet bus, although the time-division-multiplexed (TDMA) channel capacity might be as much as 10 GBits/s and the maximum propagation time would be about one nanosecond. If greater capacities are required, the optical bandwidth of the medium may be exploited using wavelength multiplexing. Multiple channels might operate in parallel at different wavelengths or wavelength routing could be used with tunable optical sources or detectors.

For multi-layer systems, it will be necessary to move data on or off of the single wafer discussed here. This can be done using point-to-point communications links between corresponding elements on adjacent wafers. The resulting system completely interconnects a high density three-dimensional array, such as that of reference[4], without the need for multiple passes which increase latency and add traffic to the network.

References

[1] J.W. Goodman, F.J. Leonberger, S-Y Kung, and R.A. Athale, "Optical Interconnections for VLSI Systems," *Proceedings of the IEEE*, Vol. 72, pp.850-866(1984).

[2] R.A. Linke, "A Proposed Optical-Interconnection Technique Employing a Two-Dimensional Waveguide as a Broadcast Medium," *Topical Meeting on Photonic Switching*, PS'91, Salt Lake City, March 6-8, 1991

[3] C.A. Brackett, "Dense Wavelength Division Multiplexed Networks: Principles and Applications," *IEEE J. on Selected Areas in Commun.*, vol. 8, no.6, pp.948-964, 1990.

[4] I. Hayashi, "Photoelectronic Integration for the Optical Computing," OC'90, Kobe, paper 9F2, 1990.

Roles of Optics in High-performance Computing Systems

Ahmed Louri
Department of Electrical and Computer Engineering
The University of Arizona
Tucson, Arizona 85721
phone : (602) 621 - 2318

In this paper, we assess the roles of optics in future computer designs. First an overview of major proposals aimed at significantly increasing computer performance is given. This is followed by a discussion on the near-term (evolutionary) role of optics in these high performance computers. We then assess the long-term (revolutionary) role of optics in future parallel computing paradigms and execution models.

A. Major Proposals for the Design of High-performance Computers

There are basically three schools of thought as to what is the most important factor in obtaining significantly higher performance. The first school believes that system speed may be increased by faster circuit and packaging technologies. This approach exploits coarse-grain parallelism by employing few (4-to-16) very complex, and possibly heterogeneous, interconnected processing elements (PEs). These PEs are supposed to operate at very high clock rates. This is the case for CRAY, ETA, Fujitsu, and Hitachi line of computers. The second school of thought puts priority on medium to fine-grain parallelism where a large number (100 to millions) of relatively small, homogeneous PEs are interconnected together. These PEs can simultaneously execute the same instruction on different data (SIMD systems such as array processors, systolic processors, content-addressable processors), or autonomously execute diverse instructions on different data (MIMD). This approach insists on retaining conventional sequential languages and architectures for the PEs and depends strongly on the use of concurrency between instructions. This concurrency must be detectable in high-level language programs and managed at the hardware level.

There are two major categories within the second school of thought, depending on the way PEs communicate. The first one is called shared-memory multiprocessors where interprocessor coordination is accomplished through a global shared-memory that each PE can address. The second one is called distributed-memory or multicomputers where several PEs, each with its own local memory, are connected with a processor-to-processor interconnection network. PEs communicate by explicitly passing messages through the interconnection network (hence the name message-passing architectures).

The third school of thought believes that a dramatic increase in performance will come from unconventional (or non-von Neumann) architectures based on new parallel models of computations that will allow dramatic exploitation of parallelism. Data-driven (dataflow) and demand-driven (reduction) computing are examples of such models. This school of thought promotes both programmability and performance. For programmability, new languages (e.g., functional languages) that are not dependent on the sequential model of computation, free from side effects, and allow explicit and implicit exploitation of concurrency are desirable. For performance, highly concurrent systems that avoid centralized control are more desirable.

B. Evolutionary Role of Optics

In the short-term, optics will complement electronics where the strength of optics lie. In the following we see how optics can help break the performance barriers faced by electronics in each major school of thought.

The first approach relies on interconnecting very powerful processors that require mass storage and a very large communication bandwidth network (Gigabits/s). However, since the number of PEs is small, optics may prove to be the ideal choice for the design of the high-speed network. In fact, a generalized crossbar would be within the capabilities of optics in this case. In addition, optical storage technology (optical disks and volume holography) may also play a fundamental role in the storage and parallel I/O requirements of such computers.

The second approach requires a large number of interconnected PEs. For the shared-memory system, the major problems are memory latency, process synchronization, and cache coherence. In principle, optics could be used to alleviate memory contention by providing contention-free parallel read access to the global memory. Moreover, the capability of broadcasting communication for global optical signals such as the clock and other synchronization signals can be used to solve the synchronization and cache coherence problems. details will be given at the workshop. The distributed-memory model relies heavily on the topology and speed of the network used to interconnect the large number of PEs. The performance of this model depends on the degree of connectivity of the communication network. Obviously, a crossbar or a fully connected network is unfeasible for this model because of the large number of PEs involved. Therefore, networks with less connectivity are usually used at the account of a longer message delivery time. However, reconfigurable optical interconnects may provide a much higher degree of connectivity at an acceptable (if not faster) message delivery time.

The most popular model for the third approach is the dataflow model. Despite the fact that the dataflow approach seems to exploit maximum parallelism, its current implementations are failing to achieve the proclaimed performance due primarily to (1) the lack of adequate communication support to satisfy the high data traffic between PEs, and (2) to the runtime overhead required to manage the tag operations and the relatively costly associative mechanism needed to implement the matching store. Clearly, optics can be used to solve both problems very efficiently. Optics can provide adequate communication support for dataflow. In addition, it can significantly reduce the runtime overhead incurred by tag matching operations. Matching symbols in optics can be implemented at a speed of light.

C. Long-Term Role of Optics

The long-term role of optics will be in the context of a uniform technology where information processing, communication, and storage are all in optical form. This uniform technology will require several key components which in my opinion are absent today. The unique properties of optics namely, spatial parallelism, speed, linear superposition, polarization, and non-interfering communications must be exploited at the component design level to produce fundamental building blocks which will open new horizons for computer architects. Because of the communication superiority of optical signals and the low degree of flexibility of optical systems, the first generation of all-optical computing architectures would likely to be based on communication-intensive computing models with multi-dimensional topologies, and exploiting fine-grain parallelism and decentralized control. In addition, optical compute-intensive special-purpose units (such as optical FFT units, matrix-manipulator units, and matching units) would be available for insertion into these main architectures.

Engineering (Design Trade-Off) Issues in Optoelectronic Computing

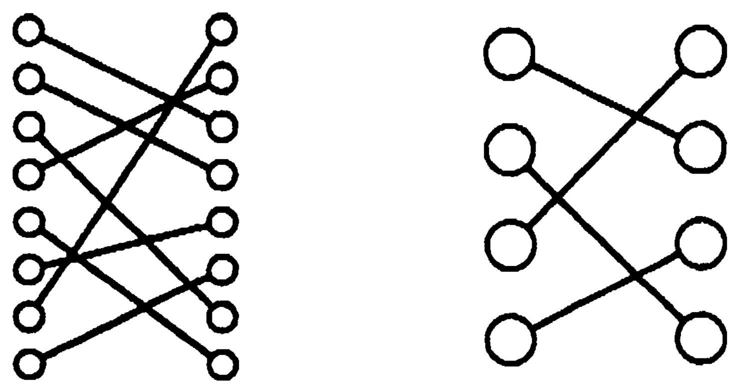
Dau-Tsuong Lu, R. Paturi, Y. Fainman, F. Kiamilev, S. Esener, and S. H. Lee
University of California at San Diego

To cope with the ever increasing demand on computing power, it is not enough to rely only on faster device technology. It is necessary to utilize parallel processing, involving a network of many processors. While the interconnections among neighboring processing elements can be done electronically, interconnections among far-away processing elements (global interconnections) can be done better optically in free space. Optoelectronic parallel computing systems utilizing free-space interconnections have been shown analytically to provide better performance (in terms of system clock speed, interconnection bandwidth and area, etc.) than pure electronic parallel computing systems when the size of the systems are scaled up.[1,2] Moreover, they can support new computing architectures (e.g. those based on expander graphs and twin butterfly interconnection networks),[3,4] which are very difficult for pure electronic systems to support.

In parallel computing there are many architectural issues to investigate. They include computational models (SIMD vs MIMD, shared vs distributed memory), I/O (memory hierarchy, auxiliary storage), interconnection network (efficient inter-processor and processor-memory communication), grain size (coarse vs fine) and fault tolerance (redundancy, reconfiguration, graceful performance degradation). At UCSD, we began to investigate the trade-off issues involved in designing interconnection networks. Based on technological constraints, we examine the trade-offs, for example, among processing elements (PE) complexity, interconnection complexity, and time. PE complexity includes the (application dependent) signal processing logic, local memory and the number of detectors and modulators (or light emitter). Optical interconnection complexity corresponds to graph complexity and hologram complexity, and is proportional to the number of detectors and modulators per PE and the array size. Time is the run time to complete an application.

A. PE Complexity vs Optical Interconnect Complexity

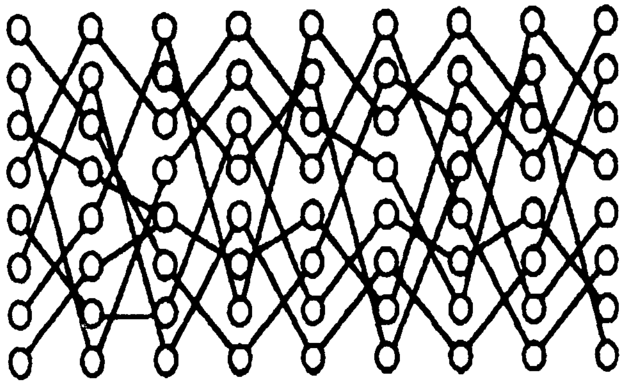
Assuming the silicon wafer and hologram areas are fixed, higher PE complexity (more signal processing logics or local memory) means smaller PE array size and reduced interconnect complexity, unless the numbers of detectors and modulators per PE are increased.

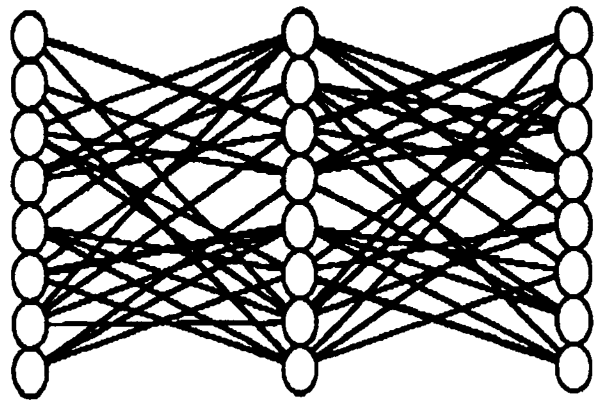


B. PE Complexity vs Time

When the number of detectors per PE is small, the PE complexity and interconnection density are also small. To implement an algorithm, a large number of interconnection stages and long reconfiguration/run time will be necessary. By increasing the number of detectors per PE, the PE complexity is increased. But, the number of interconnection stages and reconfiguration/run time can be reduced at the expenses of the interconnection density as well as hologram complexity.

Routing time on twin butterfly can be reduced by queueing and pipelining. This corresponds to an increase in local memory size in each PE.

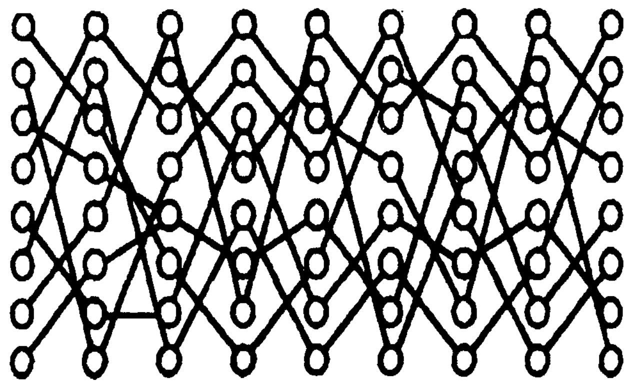


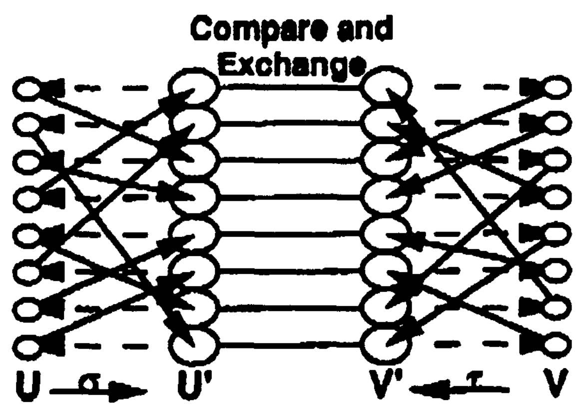


C. Optical Interconnection Complexity vs Time

We could reduce the number of interconnection stages, resulting in reduction in the interconnection complexity. However, we would need to use these limited number of stages repeatedly to achieve the original expansion, i.e. trading off time. In the following figure,the 9-stage expander graph is reduced to two stages. Data is routed from PEs in Plane U to PEs in Plane U' via random mapping σ permutation, and from PEs in Plane V to PEs in Plane V' via τ permutation. Plane U' and Plane V' will perform the compare and exchange operation using a straight-through interconnection. Another random permutation is obtained by using σ and τ twice to get $\sigma^2 \tau^2$. Thus i different one-to-one permutations are obtained in $\frac{1}{2} i^2$ time using only two interconnections.

The expander graph generation algorithm permits an interconnection between any two PEs. In the worst case, these two PEs may be located on opposite corners, resulting in a large deflection angle. By reducing the spatial randomness of these interconnections, we reduce the worst case separation of the interconnected pair of PEs. The consequent reduction in deflection angle affords decrease in interconnection as well as hologram complexity. However, reduced randomness means reduced expansion, which has to be compensated by longer running time.





Trade-offs also exist between lateral (hologram) complexity and longitudinal complexity (system length) of the interconnection system implementation. By analyzing the interconnection matrix, we could decompose a complex hologram for interconnection into a combination of simpler holograms. Hence, lateral complexity is traded for longitudinal complexity of the optical system.

References:

1. M. Feldman, S. Esener, C. Guest, and S. H. Lee, "Comparison Between Optical and Electrical Interconnects Based on Power and Speed Considerations," *Applied Optics* 27: 9, 1742-1751 (1988).

2. F. Kiamilev, S. Esener, and S. H. Lee, "Programmable Optoelectronic Multi-processor System," Proceedings *SPIE Critical Reviews of Optical Science and Technology* CR35, 197, Los Angeles, January 1990.

3. R. Paturi, D. -T. Lu, J. Ford, S. Esener, S. H. Lee, "Parallel Algorithms Based on Expander Graphs for Optical Computing," to appear in *Applied Optics* 30: 5 (1991).

4. T. Leighton and B. Maggs, "Expanders Might be Practical: Fast Algorithms for Routing Around Faults on Multibutterflies," *Proc. IEEE Symp. Foundation of Comp. Sci.* 384-389, 1989.

A Few Perspectives on Digital Optical Computing

Miles Murdocca
Department of Computer Science
Rutgers University, Hill Center
New Brunswick, NJ 08903
(908) 932-2654
murdocca@cs.rutgers.edu

Perspective #1 - Examples of how computer architecture affects optical computing
A common observation in electronic computing is that approximately 90% of the execution time of a computer program is spent in just 10% of the code, and further, that accesses to main memory by a program tend to be clustered in local areas. Thus, the use of cache memory is motivated, where just the localized part of the 10% is stored in a small, fast memory that is logically closer to the central processing unit (CPU) than is the main memory. If we make the rest of main memory as fast as the cache, then performance will not be significantly affected since most of the program is spent in just a small part of the code, so that there is little motivation for making all of computer memory out of the fast cache logic devices. This paradigm reappears in computing in many places, and if we extend this paradigm to the optical computing world, we should consider that an entire digital optical computer does not need to be made up of fast optical logic devices in order to be effective. Rather, a mix of optical/electronic or fast optical/slow optical logic devices should be considered. As an example of the latter, FLC devices are relatively slow when compared with MQW devices, but require less optical energy to switch and hold their state. Thus, FLC devices may be a good complement for MQW devices, possibly in main memory or in reconfiguring the interconnects.

Perspective #2 - In order to construct an optical computer, we have to give up some of the flexibility that we enjoy in electronic technologies.
Our work at Rutgers extends from the work done at Bell Labs. The basis of the work is an all-optical digital computer that is composed of optical logic arrays interconnected in free space. We have found that there are a lot of freedoms afforded to us in electronic technologies that we don't need in digital optical computing. For example, we can get away with regular interconnection patterns at the gate level such as perfect shuffles between optical logic gates. We can get away with fan-ins and fan-outs of only two. All of the logic on an optical logic array can be of the same type such as AND or OR, rather than an arbitrary mix of logic gates that we use in electronics. We can construct our entire computer with non-associative logic such as NOR, which means that relative logical inversions cannot be made without resorting to an alternate logic system such as dual-rail logic. We can maintain a strict logic-interconnect-repeat architecture so that all signals travel through the same number of identical logic gates. We can have all of our logic gates running at the same speeds, unlike electronic VLSI where we can take advantage of transistor sizing to trade speed for area. We can live with all of these restrictions in the interest of simplifying the construction of optical processors, but when all of these restrictions are taken in conjunction, performance suffers significantly. The primary areas of cost that these restrictions affect are gate count and circuit latency. The suggestion here is that something has got to improve, for example fan-in and fan-out, or maybe the complexity of the interconnects.

Perspective #3 - Symbolic substitution

I joined Alan Huang's group at Bell Labs in 1983, which was the year that Alan presented his symbolic substition paper at the International Optical Computing Conference in Cambridge. At the time, Alan explained that he was trying to get people to work in the image planes rather than in the Fourier domain, and to show that an optical computer only needs simple configurations of optical logic arrays with simple, regular interconnection patterns. At Bell Labs we explored this area for a while, and it paid off in a number of ways. For example, device people were encouraged to continue working on optical logic arrays, optical systems people started looking into implementation problems, and architecture people looked into the problem of mapping arbitrary problems onto the regularized model. As work progressed, people realized that a more efficient computer can be constructed if we treat the optical logic gates as logic gates, and the interconnects as interconnects, rather than mapping problems into symbolic substitution first. My opinion is that symbolic substitution has had a great influence in advancing some areas of optical computing, and that it is certainly academically interesting, but that it is not very practical when compared with more direct methods. The basic model of logic-interconnect-repeat that a number of researchers use today hasn't changed from Huang's original proposal, but the architectures are different since the emphasis is no longer on symbolic substitution.

Perspective #4 - Can optics do something that electronics cannot?

It appears that optical logic gates cannot operate faster than electronic logic gates, because it is the same underlying phenomena that governs switching in optics and electronics. There is a 250 MHz electronic RISC processor that has been demonstrated at Rensselaer. Given that there is no fundamental reason that optical switching should be faster than electronic switching, and that electronic switching is already so fast, the question arises as to whether there is something that optics can do that electronics cannot, or is it just a matter of better engineering?

Perspective #1 above hints that not all of a computer needs to be optical in order to appreciate a gain in performance. Perspective #4 (this one) argues for an all-optical digital computer because there are some profound things that an all-optical technology can do that an electronic or hybrid optical/electronic technology cannot do. When all of the gate-level interconnections are in free space, then we know that we can only have faults in the active logic gates. Further, when all of the optical logic gates are isolated from each other, as in S-SEED arrays and other optical logic arrays, then we can independently observe each logic gate for failure, and isolate failed logic gates from others by performing some manipulation in free space such as a masking operation. Thus, we can deal with greater chip sizes and poorer yields than electronic technologies allow. We also have the capability to completely rewire the gate-level interconnect of an optical computer on every time step. So far example, we can have a 68000 processor on one time step, a SPARC processor on the next, and a signal processor on yet another time step. At Rutgers we are exploring the development of a compiler that generates object code for an architecture that it also produces. This is a profound departure from conventional electronic computing, and obviously cannot be supported by conventional electronic computing as long as we use physical wires to carry information. The question that remains to be answered is how important these issues are.

Fine-grain parallelism: A Simple Machine

Michael T. Pope
AT&T Bell Laboratories

Two key areas of interest in parallel computer architectures are---

- Systems to handle problems with an extremely high degree of parallelism
- Automatic extraction of parallelism inherent in a problem

To achieve high levels of parallelism, processing elements must be numerous, and therefore, simple. We present an architecture consisting of a linearly connected "stream" of aggressively simplified processors and local memory elements.

The processors essentially implement a variation of Turner's combinators--- these are "string rewriting" transformations of little computational complexity. Combinators are an execution technique proposed for functional programming languages, and have the desirable property of automatically exposing data-movement parallelism. The combinators flow down the stream, and are acted upon opportunistically by the processors as they pass. Eventually no further reductions occur, leaving the results in the stream.

Additional "pseudo-combinators" can be added for non-data-movement purposes--- for example arithmetic or list processing. Computationally expensive pseudo-combinators (for example, division) can be relegated to special pseudo-combinator specific processors distributed relatively infrequently throughout the stream. This scheme is highly flexible, potentially allowing streams to be optimized for specific problems, but there is potential for latency problems--- this is an area of ongoing refinement, as the tradeoffs in placement and coalescing of different processor types are currently unclear.

This architecture is currently being investigated with simulations in preparation for VLSI implementation later in 1991. Illustrations drawn from simulations of one detailed potential hardware partitioning of this architecture are presented.

Optical Computers or Optics in Computing?

Michael Prise
AT&T Bell Labs.
Rm 4G-526
Crawford's Corner Rd.
Holmdel, NJ 07733

In this paper I will address computer architects as an abstract group (not including me!). In Webster's dictionary an architect is defined as
1. One who designs (computers) and advises in their construction.
2. One who plans and achieves a difficult objective.
Most of us attending this workshop can be described as potential computer architects since we have offered plenty of advice in the form of papers and talks. Also the architecture must take into account the devices, and the packaging, as well as the overall system layout so it is impossible to separate these aspects. This makes understanding and communication between experts in these different fields essential to the overall system architecture. Which means we have to have workshops like this!

We at AT&T, as well as several other groups have put together a number of demonstration systems which we have all somewhat arbitrarily designated as optical processors (computers?). My working definition of an optical processor is one in which all the connections between logic gates are optical. The devices we are using have also somewhat arbitrarily been defined as optical logic gates, because the logical inputs and outputs are optical. Our particular devices are more clearly described as electro-optic. There is no such thing as an all-optical gate or an all-optical computer. Since the system definitions are so unclear it is natural to ask what is meant by optical computing architectures, and in particular what if anything is different from conventional computer architectures? Before trying to answer this question I would like to give my own entirely unbalanced view of the status of optics in computing.

Optical fiber systems are clearly useful for longer distance communications, and are finding increasing uses for shorter distances. Systems operating at 200Mbit/s over 100s of feet are becoming commercially viable. Fiber systems operating over distances of 10s of feet and operating at >500Mbits/s will probably become viable in the near term.

Many proposals and simple demonstrations have been carried out using both free space and waveguide systems to implement optical interconnects at the backplane level, but it is unclear when or whether any of these are going to become viable in real machines.

As I stated in the first paragraph, we have defined an optical processor (computer) to be one which uses optical interconnect at the gate-to-gate level. What I did not say is that the technologies we have developed can be used to provide optical interconnect at the chip to chip level. We can use the same device technology to provide optical I/O with some electrical processing. Indeed we could view the end result as a VLSI chip with optical input and output, the exact partitioning depending on the architectural design. The viability of this approach is also unclear, although good physical justifications in terms of power dissipation would encourage us to believe that eventually its time will come.

The sequence in which I have presented my views represents an evolutionary view, with optics slowly penetrating the communication hierarchy of the computer. It implies no astounding leaps in architecture. It would represent a gradual evolution from an all electronic system, as the amount of optics used increases, and would not necessitate a particular field called "Optical Computing Architecture".

However if the latter approach, with optical interconnect at the chip to chip and gate to gate level becomes technically feasible, before the intermediate approach of waveguide or fiber backplane connectors, this would leave the door open for revolutionary architectural approaches, with systems consisting of chips with dense, fast and regular optical I/O. Designing systems using this type of technology should provide some challenges for would-be optical computer architects.

So far, many advances have taken place in the device technology at least at the research level, and we have put together some very primitive systems using these devices. These systems have proven that given enough space and money we can build "optical processors". In order to get to the next stage - beyond research - we need a serious system drive to generate sufficient resources and to find out what the real problems are. Telecommunications switching systems seem to be our principle drive just now. These systems tend to have much higher communications requirements than computers, making the use of optics in the short term more likely.

The challenge for optical computer architects is to develop some more specific system goals and some specific architectures so we can continue working on new device and packaging technologies. We need some focus in order to attack the real systems issues involved in implementing a particular design. If we have no focus optical interconnect is going to penetrate in an evolutionary manner and there will be no field of optical computer architecture, it'll just be computer architecture. Similarly the field of optical computing will simply be engulfed by the field of high speed digital system design.

My own view, which I hope will change during this conference, is that the evolutionary approach is more likely to prevail. This would entail taking a system being designed using digital electronics and seeing if optics can offer some performance enhancement. In this way we are leveraging off previous system design. Perhaps then then we can develop sufficient technologies that novel architectural approaches will become viable. I hope somebody comes up with a novel architecture using optics which is demonstrably better than any conventional architectures, since I would much rather participate in a revolution than evolution.

Approaching the problem of high density interconnects in the 'Monsoon' data-flow parallel processor with optics.

Fred Richard and Michael Lebby

Motorola Inc.
Phoenix Corporate Research Laboratories
Photonics Technology Center

The 'Monsoon' data-flow parallel processing project has evolved initially through funding by DARPA to MIT. Presently, to satisfy the industrial partner requirements, Motorola Inc., has taken the responsibility to demonstrate a working 16 node processor version in 1991. It is becoming evident that the interconnect problem for such systems may not be easily solved using electrical interconnect designs common in the computer industry today. New, novel optical methods may be required that will offer both cost, and performance advantages over existing electrical approaches.

The essence of this project is not to directly find solutions to the 'optical computer' problem that has been researched extensively over the last two decades, but to examine the role of optical interconnects and finding a competitive technology that can co-exist with or even replace the current electrical technologies.

It has been argued that a general purpose computer consisting of multiple processors must be scalable, i.e., one can show that when the node number is increased, the performance will increase. It has also been stated that the architectures of such systems must address the system level inefficiency problem which turns out to be a result of both memory latency and idling due to synchronization requirements. If a well engineered parallel processor is constructed, these inefficiencies could be reduced through the sheer parallelism of the program. Latency, could be improved simply by pipelining the instructions, but the limitations of single thread computation in a Von Neumann processor makes this solution temporary. Other techniques that could be used to reduce latency unfortunately increase the burden on switching capabilities. It has also been argued that the cost of synchronization in serial Von Neumann architectures is prohibitive. Data-flow systems like Monsoon actually treat each instruction as a task and by allowing very small hardware synchronization cost per executed instruction, offer excellent flexibility in scheduling instructions to reduce processor idle time.

In data-flow computing the processor performs the computation as soon as all the necessary input data is available, or waits until the results of the computation are demanded by other processors. The importance of this technique is that data flows through the system in parallel and provides a high level of concurrency in computation. However, it has one big disadvantage that has prevented it from becoming a practical success: the enormous requirement of interconnections. For the case of a 16 node system the need for an optical interconnect technology is questionable, but for larger systems with many more nodes interfacing with each other, the case for optics becomes much stronger. The Monsoon switching system routes data packets to and from processing elements directly using 4X4 cross-bar switches as shown in Figure 1. Each processing element has dedicated input and output ports each to a 4X4 switch. Monsoon's performance depends on the ability to send and receive data reliably, and presently the data link ASIC's do not have error correction capability (which adds complexity for an optical interconnect solution). The current design goal is for a 500 hour continuous operation between errors and at a data transmission rate of 800Mbps (point to point). The bit error calculates to be 7x10E-18 failures/bit. The data link system will operate on a 200 Mhz clock.

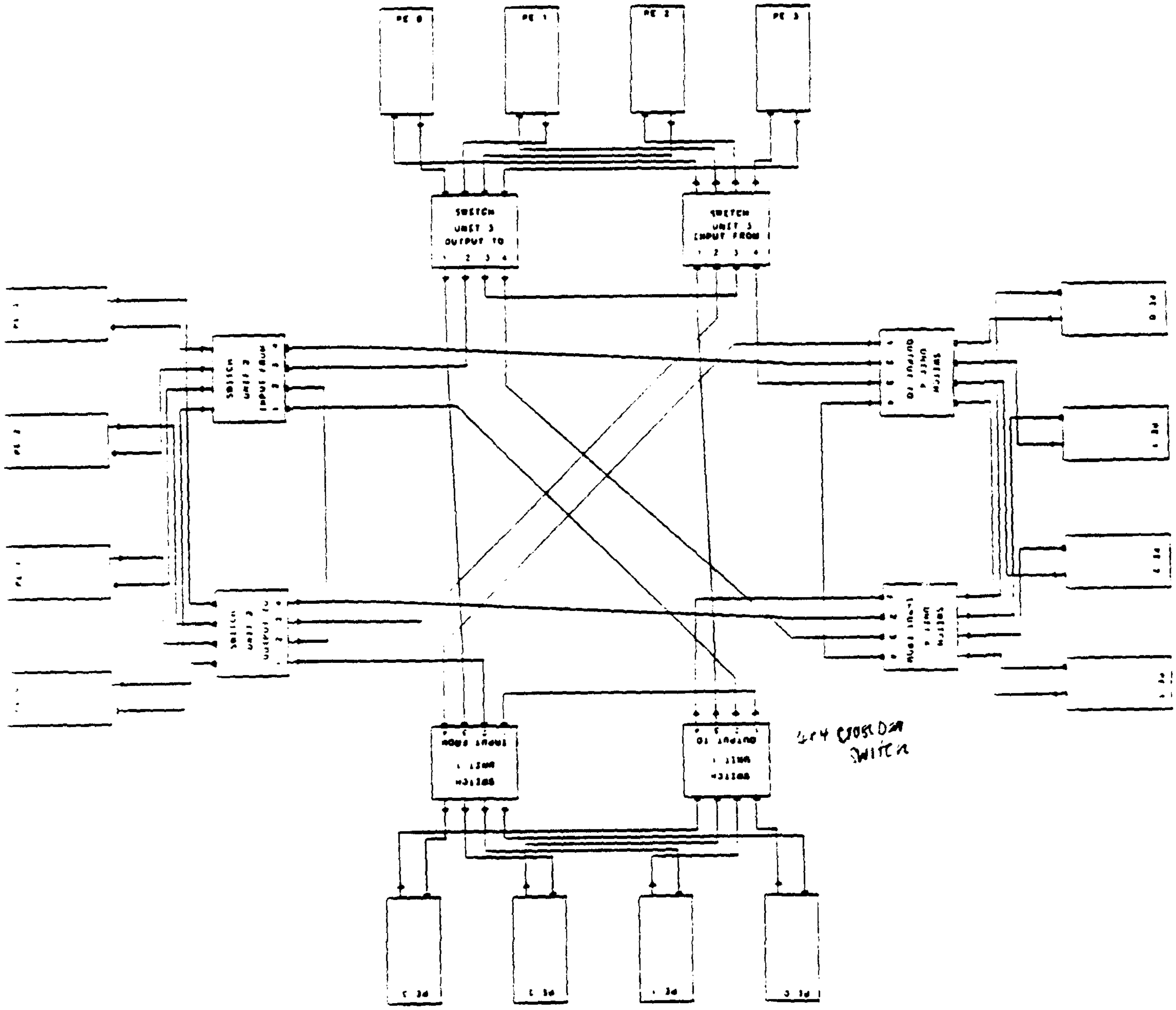
The role of optics in data-flow systems has been discussed recently by Louri [1] who detailed a scheme for an optical data flow computer. Louri discusses the mapping of fine-grain data-flow architectures using a limited fan-in and fan-out methodology. Hardware complexity is cited as the road-block to pure data-flow processing with respect to a processing plane. Free space node to node optical interconnects are suggested with appropriately defined message protocols. Although this approach may eventually become the back-bone for the data-flow processor, the in-plane approach to optical interconnects using polymer waveguides or holographic materials may be the more near term approach with respect to the cost and manufacturing requirements.

The current optical interconnect question for such a data-flow project is deciding which of the two dominant approaches (free-space or in-plane) to follow. The free-space technique offers vastly different architecture designs with allowances for more exotic routing; higher integration densities and longer interconnect distances. Unfortunately, the cost of the components for such systems with the stringent requirements for vibration and alignment may make this solution more long-term. As the problems of cross-talk are eliminated, the third dimension allows communication between the planes of chip instead of the edges and therefore permits a significant level of parallelism over the two dimensional approach. A dominant issue that will need additional addressing is how to configure and reconfigure the many beams of light and focus them into the correct ports at data rates close to 1Gbps and very low bit error rates (as per the Monsoon project). Holographic elements may become one manufacturable solution, or if lenslet arrays could be fabricated with adequate precision, then this may become a viable alternative.

The in-plane arguments usually include easier alignment tolerances and easier manufacturability with the potential for significant economies of scale. In addition, this technology allows a better transition from the electrical domain to the optics domain even though the media utilised may be completely different. Although the in-plane argument is not considered a true solution in the third dimension, it does however, present an incognito 3D solution. One interesting area that may fit into this category are the substrate mode holograms. Here, the optical medium can be used to focus light to predetermined receivers while still allowing beams to cross without interference. The polyimide approach has been the most developed to date, but the use of formed waveguides that use the air interface for cornering may restrict interconnect densities if multi-layer designs are required for large data-flow systems. The issue of power consumption for the in-plane solutions may be one draw-back for very large systems with many interconnect links. Component reliability in hostile operating conditions may well be a major element in the decision on performance specifications for data-flow computers.

In summary, for very low bit error rate levels as required by the Monsoon project, component redundancy, reliability, manufacturability, performance and cost will become influential in the decision to pursue a particular optical interconnect technology.

[1] Louri, A., "An Optical Data-flow Computer," SPIE Vol 1151 Optical Information Processing Systems and Architectures (1989), pp47-58



16 PE NETWORK
FIGURE 1

The Fundamental Limit of the Reliability of Optical Logic

Charlie Stirk

January 24, 1991

1 Problem Statement

What is the fundamental quantum limit on the reliability of an ideal optical logic device in the presence of shot noise?

2 Background

A sum-and-threshold optical logic device has a step function response to the intensity of light. Recall that the intensity of light is Poisson distributed

$$\Pr(\text{k-events in n-tries when probability is } \lambda) = P_{\lambda n}(k) = \frac{(\lambda n)^k}{k!} e^{-\lambda n} \quad (1)$$

where λn is the mean number of detected photons. The logic is ideal when the mean number of photons for a logic low is zero. The bit-error-rate (BER) is the probability that the device will produce an erroneous output when the symbols on the input channels are independent and equiprobable.

2.1 Optical OR

For the optical OR gate with fan-in N, the BER at threshold T is

$$BER_{OR}(T) = \sum_{i=0}^{N-1} \binom{N}{i} \left(\frac{1}{2}\right)^N \int_0^T P_{im_L+(N-i)m_H}(k)dk + \left(\frac{1}{2}\right)^N \int_T^\infty P_{Nm_L}(k)dk \quad (2)$$

where m_L and m_H are the mean number of photons for logic low and high, respectively. The first term is the probability that an output high is miss-classified as a low and the second term is the reverse situation. Since the logic is ideal, $m_L = 0$, and thus, the expression reduces to

$$BER_{OR}(T) = \sum_{i=1}^{N} \binom{N}{i} \left(\frac{1}{2}\right)^{N} \int_{0}^{T} P_{im_H}(k)dk \tag{3}$$

If the threshold is zero, everything gets classified as a logic high, producing a BER of $\frac{1}{2}$. Thus. the lowest BER for the optical OR occurs when the threshold is equal to one, which minimizes the integral. Using equation 1 when $k = 0$ in equation 3 gives us the fundamental quantum limit on the BER (FBER) of an optical OR with fan-in N.

$$FBER_{OR} = \sum_{i=1}^{N} \binom{N}{i} \left(\frac{1}{2}\right)^{N} \exp^{-im_H} \tag{4}$$

If we expand this in terms of N

$$FBER_{OR} = \left(\frac{1}{2}\right)^{N} \left[N \exp^{-m_H} + \frac{N(N-1)}{2} \exp^{-2m_H} + \ldots + \exp^{-Nm_H}\right] \tag{5}$$

Because of the exponential, for sufficiently large m_H

$$FBER_{OR} \approx \left(\frac{1}{2}\right)^{N} N \exp^{-m_H} \tag{6}$$

Note that when $N = 1$, this is the fundamental quantum limit of a detector.

$$FBER_{detector} = \frac{1}{2} \exp^{-m_H} \tag{7}$$

Equation 6 is plotted in figure 1. Note that the number of photons per bit needed to obtain a given BER decreases slightly with increasing fan-in and is just barely better than a detector.

Equation 4 is plotted in figure 2. Without the approximation of equation 6, the number of photons required for a given BER is a little more than that of figure 1 due to the extra terms in the sum.

2.2 Optical AND

For the optical AND gate with fan-in N, the BER at threshold T is

$$BER_{AND}(T) = \sum_{i=0}^{N-1} \binom{N}{i} \left(\frac{1}{2}\right)^{N} \int_{T}^{\infty} P_{im_H+(N-i)m_L}(k)dk + \left(\frac{1}{2}\right)^{N} \int_{0}^{T} P_{Nm_H}(k)dk \tag{8}$$

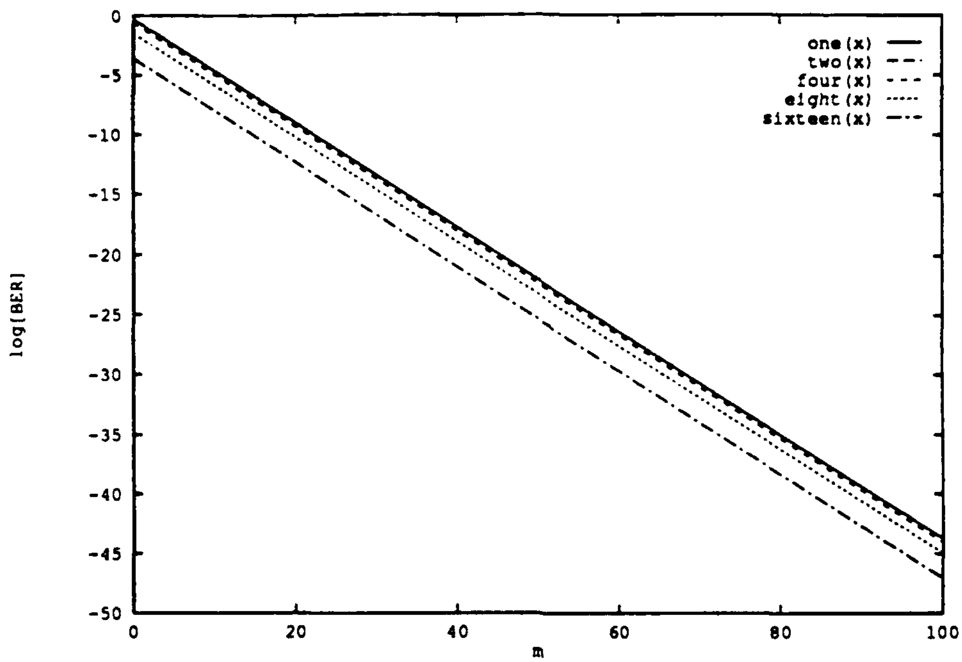


Figure 1: Approximate quantum limit on log[BER] of an optical OR as given by equation (6) vs mean number of photons per channel for a logic high for several fan-ins.

where m_L and m_H are the mean number of photons for logic low and high, respectively. The first term is the probability that an output low is miss-classified as a high and the second term is the reverse situation. For an ideal device ($m_L = 0$) this reduces to

$$BER_{AND}(T) = \sum_{i=0}^{N-1} \binom{N}{i} \left(\frac{1}{2}\right)^N \int_T^\infty P_{im_H}(k)dk + \left(\frac{1}{2}\right)^N \int_0^T P_{Nm_H}(k)dk \qquad (9)$$

When this is minimized with respect to T we get the FBER for a given N and m_H. This is shown in figure 3.

In contrast to what we found for the optical OR, for the optical AND the minimum number of photons necessary for a given BER increases with fan-in.

3 Conclusions

The fundamental quantum limit on the BER of an ideal sum-and-threshold optical logic device is determined by shot noise. This class of optical devices includes conventional detectors, surface emitting laser diodes, vstep and nonlinear Fabry-Perot etalons. It does not include differential devices such as the seed or more exotic devices like soliton switches. NOR and NAND devices have

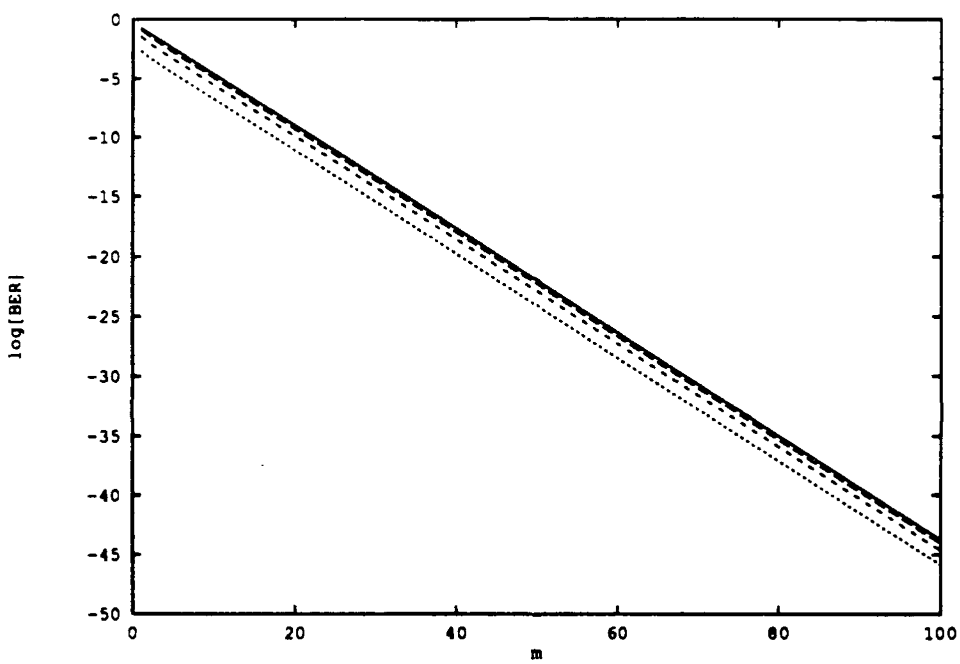


Figure 2: Exact quantum limit on log[BER] of an optical OR as given by equation (4) vs mean number of photons per channel for a logic high for several fan-ins.

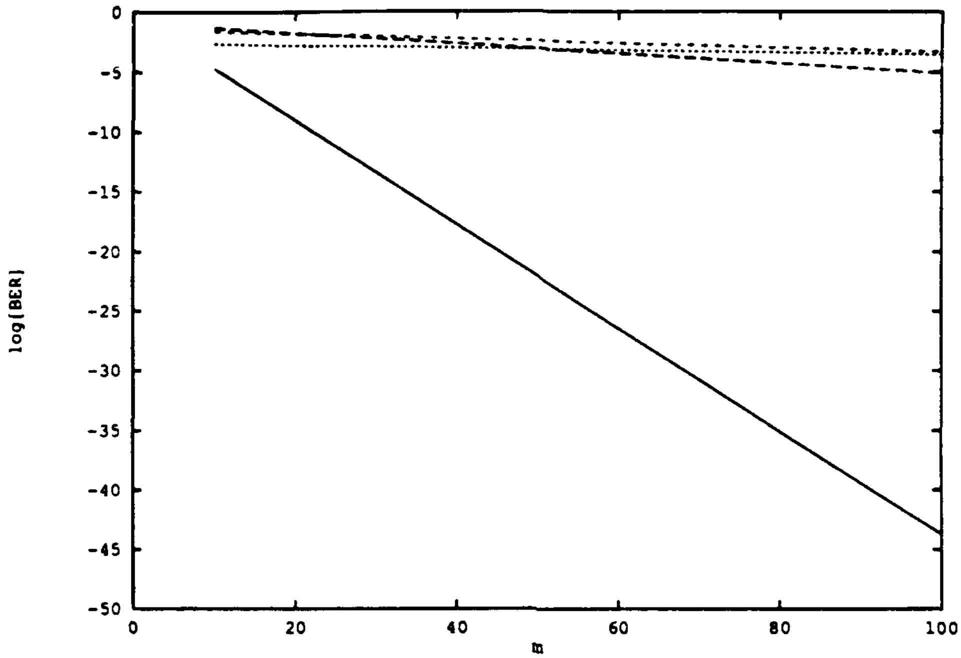


Figure 3: Quantum limit on log[BER] of an optical AND as given by the minimum of equation (9) with respect to T vs mean number of photons per channel for several different fan-ins.

the same reliability characteristics as the OR and AND devices, respectively. The main result of this exercize is that optical AND's should only be used at very low fan-ins to ensure reliable operation.

Asynchronous and fault-tolerant design in digital optical computing

Kelvin Wagner

Optoelectronic Computing Systems Center
Department of Electrical and Computer Engineering
University of Colorado at Boulder

In many of the envisioned applications of optical switching systems, the arrival time of signals and control operators can not be guaranteed to be synchronous, or even vaguely coincident. An example is a telecommunication switching system, where the various inputs are from geographically distributed sources, from which it is impossible to guarantee synchronous arrivals. This manifestly requires an asynchronous communication protocol to be used, which necessitates the incorporation of speed independent circuits, and forbids the use of clocks. This may eliminate the possibility of time multiplexed gain as utilized by S-SEED circuits, and the soliton dragging switch, and fundamentally requires some type of latching behavior. The conventional latch used is self-timed VLSI systems is the Mueller C-element which is equivalent to a majority gate with feedback. The natural threshold logic implementation of a majority gate immediately suggests the possibility of implementing a C-element using optical bistability. Various types of optical bistable devices can be considered, including nonlinear etalons, SEEDs, and microlasers, but increasing absorption based systems with clockwise loops are not directly applicable. A bias beam which is below the down switch threshold, $I_\downarrow$, is applied to the optical bistable device, so that with no input the output is always low $Bias + 2Low < I_\downarrow$, where Low is the output of the device when it is off. This beam is never clocked, the self-timed signals themselves clock the device. When only one of the two inputs goes high, the device is biased into the middle of the bistable loop, but does not switch on, so the output remains low, $I_\downarrow < Bias + Low + High < I_\uparrow$, where $High$ is the on state output, and $I_\uparrow$ is the switch on threshold. Upon application of both inputs, the device is biased above the switch up threshold, $I_\uparrow < Bias + 2High$, and the device switches into the high state. When one of the two inputs is removed, the device remains in the middle of the bistable loop, so the output remains high. It does not switch back to the low state until both inputs are removed, beginning another cycle. the tolerancing of this mode of operation is not much worse than a 2 input AND gate, but a wide bistable loop is required. This application provides a key motivation to pursue optical bistability for asynchronous digital optical computing architectures, and is a contradiction of the widely repeated statement that bistability is neither wanted nor needed.

Another key requirement in a practical digital optical computing scheme is the ability to recover from transient and permanent device errors. Redundancy is the most common technique to endow a system with limited fault tolerance and increase the system reliability beyond that given by the product of the probabilities of correct operation of the components. However, in most redundant systems, a voter is required to resolve conflicts between the redundant components, but a fault in the voter still produces erroneous outputs. Although multiply redundant voters can be incorporated, a mechanism must be included that eliminates and replaces faulty elements from the circuits, or else errors can propagate, but these

techniques can become quite complex and may be inappropriate for optical implementation. Another approach is to distribute the voter throughout the circuit using the technique of quadded logic. In quadded logic, 4 copies of the circuit are produced, then interconnected in a permuted fashion that allows single errors within a small block of elements to be detected and corrected. This is an expensive approach to fault tolerance, since it multiplies the hardware by a factor of 4, and doubles the fanout and fanin of the elements. The optical implementation of quadded logic in a regularly interconnected split-shift-mask system has some attractive features. The depth of the circuit is not increased, just the width, so no additional delay and speed penalties are imposed. The interconnections between the quadded circuits are quite regular and may be amenable to optical implementations. The even layers have a duplication of the original interconnection topology, plus an additional fanout to a cyclically permuted subset of the quadded gates in the next layer. The odd layers fanout their outputs in the same topology as the original circuit but with a cyclical magnification of two. These interconnection topologies are reminiscent of the split-shift approaches to shuffles and crossovers, and might be implemented with a similar technique. An attractive possibility is to interleave the original circuits layers on rows separated by 4, and interleave the quadded duplicates on the intervening layers. The same basic architecture of split-shift-mask-shuffle within the rows can be performed, and holographic interconnections within a quadded set of 4 rows might not overly increase the system complexity. This approach to redundant fault tolerant digital optical computing may allow the utilization of devices with increased probabilities of failure without a system reliability penalty. Such an approach may be required in order to make practical and reliable digital optical computers.

List of attendees

Dan Blumenthal
Optoelectronic Computing Systems Center
University of Colorado
Campus Box 525
Boulder, CO 80309-0525
danb@fred.colorado.edu

Julian Bristow
Honeywell, Inc.
Systems and Research Center
10701 Lyndale Ave. South
Bloomington, MN 55420-5601
(612) 887-4023

David Casasent
Carnegie Mellon University
Ctr. for Excellence in Optical Data Proc.
5000 Forbes Avenue
Pittsburgh, PA 15213-3890
(412) 268-2464

Pierre Chavel
Centre Univ. d'Orsay
Institut d'Optique Theorique et
Appliquee/CNRS
BP 3
F-91406 Orsay Cedex, France

Julian Cheng
University of New Mexico
EECE Building
Albuquerque, NM 87131
(505) 277-5605

Thomas Cloonan
AT&T Bell Labs, Room 2F-233
200 Park Plaza
P.O. Box 3050
Naperville, IL 60566-7050
(312) 713-5441
att!iexist!cloonan

Alan Craig
Air Force Office of Scientific Research
Building 410
Bolling AFB, DC 20332-6448
(202) 767-6337
craig@ccf2.nrl.navy.mil

Alex Dickinson
AT&T Bell Labs, Room 4E-514
Crawfords Corner Road
Holmdel, NJ 07733
(201) 949-1083
att!vax135!alex

Timothy Drabik
Georgia Institute of Technology
Sch. Elec. Eng.
Atlanta, GA 30332
(404) 853-9449

Sadik Esener
Univ. of California/San Diego
Dept. of Elec. and Comp. Eng.
La Jolla, CA 92093
(619) 534-2732
sadik@celece.ucsd.edu

C. Lee Giles
NEC Research Institute
4 Independence Way
Princeton, NJ 08540
(609) 951-2642
giles@research.nec.com

Michael Haney
BDM International, Inc.
7915 Jones Branch Drive
McLean, VA 22102
(703) 848-5854

Alan Huang
AT&T Bell Labs, Room 4G-514
Crawfords Corner Road
Holmdel, NJ 07733
(201) 949-7631
att!vax135!alan

Fouad Kiamilev
Univ. of California/San Diego
Dept. of Elec. and Comp. Eng.
La Jolla, CA 92093

Michael Lebby
Motorola, Inc.
2100 E. Elliot Road, M/S EL508
Tempe, AZ 85284

Michael Pope
c/o Joan Fauci
AT&T Bell Labs, Room 4F-538
Crawfords Corner Road
Holmdel, NJ 07733
(201) 949-3004 (sec.)

Sing Lee
Univ. of California/San Diego
Department of Elec. and Comp. Eng.
La Jolla, CA 92093
(619) 534-2413
lee@celece.ucsd.edu

Kelvin Wagner
Univ. of Colorado/Boulder
Box 425, Dept. of ECE
Boulder, CO 80309-0425
(303) 492-4661
kelvin@turkish.colorado.edu

Yao Li
Department of Electrical Engineering
The City College of the City University of New York
Convent Ave. at 138th St.
New York, NY 10031
(212) 690-6960

Richard Linke
NEC Research Institute
4 Independence Way
Princeton, NJ 08540
(609) 951-2632
ral@research.nec.com

Anmed Louri
Dept. of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721
louri@paris.ece.arizona.edu

Miles Murdocca
Dept. of Computer Science
Rutgers University, Hill Center
New Brunswick, NJ 08903
(908) 932-2654
murdocca@cs.rutgers.edu

Michael Prise
AT&T Bell Labs, Room 4G-526
Crawfords Corner Road
Holmdel, NJ 07733
(908) 949-8716
att!vax135!mep

Priscilla Rasmussen (Workshop Admin.)
Dept. of Computer Science/LCSR
Rutgers University, Hill Center
New Brunswick, NJ 08903
(908) 932-2768
rasmussen@cs.rutgers.edu

Charles Stirk
California Institue of Technology
MS 116-81, Elec. Eng. Dept.
Pasadena, CA 91125
charlie@sunoptics.caltech.edu